

# The evolution of clustering length, large-scale bias and host halo mass at $2 < z < 5$ in the VIMOS Ultra Deep Survey (VUDS)\*

A. Durkalec<sup>1</sup>, O. Le Fèvre<sup>1</sup>, A. Pollo<sup>19,20</sup>, S. de la Torre<sup>1</sup>, P. Cassata<sup>1,18</sup>, B. Garilli<sup>3</sup>, V. Le Brun<sup>1</sup>, B.C. Lemaux<sup>1</sup>, D. Maccagni<sup>3</sup>, L. Pentericci<sup>4</sup>, L.A.M. Tasca<sup>1</sup>, R. Thomas<sup>1</sup>, E. Vanzella<sup>2</sup>, G. Zamorani<sup>2</sup>, E. Zucca<sup>2</sup>, R. Amorín<sup>4</sup>, S. Bardelli<sup>2</sup>, L.P. Cassarà<sup>3</sup>, M. Castellano<sup>4</sup>, A. Cimatti<sup>5</sup>, O. Cucciati<sup>5,2</sup>, A. Fontana<sup>4</sup>, M. Giavalisco<sup>13</sup>, A. Grazian<sup>4</sup>, N. P. Hathi<sup>1</sup>, O. Ilbert<sup>1</sup>, S. Paltani<sup>9</sup>, B. Ribeiro<sup>1</sup>, D. Schaerer<sup>10,8</sup>, M. Scodeggio<sup>3</sup>, V. Sommariva<sup>5,4</sup>, M. Talia<sup>5</sup>, L. Tresse<sup>1</sup>, D. Vergani<sup>6,2</sup>, P. Capak<sup>12</sup>, S. Charlot<sup>7</sup>, T. Contini<sup>8</sup>, J.G. Cuby<sup>1</sup>, J. Dunlop<sup>16</sup>, S. Fotopoulou<sup>9</sup>, A. Koekemoer<sup>17</sup>, C. López-Sanjuan<sup>11</sup>, Y. Mellier<sup>7</sup>, J. Pforr<sup>1</sup>, M. Salvato<sup>14</sup>, N. Scoville<sup>12</sup>, Y. Taniguchi<sup>15</sup>, and P.W. Wang<sup>1</sup>

<sup>1</sup> Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, 13388, Marseille, France

<sup>2</sup> INAF-Osservatorio Astronomico di Bologna, via Ranzani,1, I-40127, Bologna, Italy

<sup>3</sup> INAF-IASF, via Bassini 15, I-20133, Milano, Italy

<sup>4</sup> INAF-Osservatorio Astronomico di Roma, via di Frascati 33, I-00040, Monte Porzio Catone, Italy

<sup>5</sup> University of Bologna, Department of Physics and Astronomy (DIFA), V.le Berti Pichat, 6/2 - 40127, Bologna, Italy

<sup>6</sup> INAF-IASF Bologna, via Gobetti 101, I-40129, Bologna, Italy

<sup>7</sup> Institut d'Astrophysique de Paris, UMR7095 CNRS, Université Pierre et Marie Curie, 98 bis Boulevard Arago, 75014 Paris, France

<sup>8</sup> Institut de Recherche en Astrophysique et Planétologie - IRAP, CNRS, Université de Toulouse, UPS-OMP, 14, avenue E. Belin, F31400 Toulouse, France

<sup>9</sup> Department of Astronomy, University of Geneva, ch. d'cogia 16, CH-1290 Versoix, Switzerland

<sup>10</sup> Geneva Observatory, University of Geneva, ch. des Maillettes 51, CH-1290 Versoix, Switzerland

<sup>11</sup> Centro de Estudios de Física del Cosmos de Aragón, Teruel, Spain

<sup>12</sup> Department of Astronomy, California Institute of Technology, 1200 E. California Blvd., MC 249-17, Pasadena, CA 91125, USA

<sup>13</sup> Astronomy Department, University of Massachusetts, Amherst, MA 01003, USA

<sup>14</sup> Max-Planck-Institut für Extraterrestrische Physik, Postfach 1312, D-85741, Garching bei München, Germany

<sup>15</sup> Research Center for Space and Cosmic Evolution, Ehime University, Bunkyo-cho 2-5, Matsuyama 790-8577, Japan

<sup>16</sup> SUPA, Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, EH9 3HJ, United Kingdom

<sup>17</sup> Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

<sup>18</sup> Instituto de Física y Astronomía, Facultad de Ciencias, Universidad de Valparaíso, Av. Gran Bretaña 1111, Casilla 5030, Valparaíso, Chile

<sup>19</sup> Astronomical Observatory of the Jagiellonian University, Orla 171, 30-001 Cracow, Poland

<sup>20</sup> National Centre for Nuclear Research, ul. Hoza 69, 00-681, Warszawa, Poland

Preprint online version: November 24, 2014

## ABSTRACT

We investigate the evolution of galaxy clustering for galaxies in the redshift range  $2.0 < z < 5.0$  using the VIMOS Ultra Deep Survey (VUDS). We present the projected (real-space) two-point correlation function  $w_p(r_p)$  measured by using 3022 galaxies with robust spectroscopic redshifts in two independent fields (COSMOS and VVDS-02h) covering in total  $0.8 \text{ deg}^2$ . We quantify how the scale dependent clustering amplitude  $r_0$  changes with redshift making use of mock samples to evaluate and correct the survey selection function. Using a power-law model  $\xi(r) = (r/r_0)^{-\gamma}$  we find that the correlation function for the general population is best fit by a model with a clustering length  $r_0 = 3.95^{+0.48}_{-0.54} \text{ h}^{-1} \text{ Mpc}$  and slope  $\gamma = 1.8^{+0.02}_{-0.06}$  at  $z \sim 2.5$ ,  $r_0 = 4.35 \pm 0.60 \text{ h}^{-1} \text{ Mpc}$  and  $\gamma = 1.6^{+0.12}_{-0.13}$  at  $z \sim 3.5$ . We use these clustering parameters to derive the large-scale linear galaxy bias  $b_L^{PL}$ , between galaxies and dark matter. We find  $b_L^{PL} = 2.68 \pm 0.22$  at redshift  $z \sim 3$  (assuming  $\sigma_8 = 0.8$ ), significantly higher than found at intermediate and low redshifts. We fit an HOD model to the data and we obtain that the average halo mass at redshift  $z \sim 3$  is  $M_h = 10^{11.75 \pm 0.23} \text{ h}^{-1} \text{ M}_\odot$ . From this fit we confirm that the large-scale linear galaxy bias is relatively high at  $b_L^{HOD} = 2.82 \pm 0.27$ . Comparing these measurements with similar measurements at lower redshifts we infer that the star-forming population of galaxies at  $z \sim 3$  should evolve into the massive and bright ( $M_r < -21.5$ ) galaxy population which typically occupy haloes of mass  $\langle M_h \rangle = 10^{13.9} \text{ h}^{-1} \text{ M}_\odot$  at redshift  $z = 0$ .

**Key words.** Cosmology: observations – large-scale structure of Universe – Galaxies: high-redshift – Galaxies: clustering

## 1. Introduction

Since the very first galaxy surveys have revealed the complex structure of the Universe, the mapping of its evolution has become an important part of Cosmology. Large surveys revealed that the Universe is composed of dense regions like clusters and filaments, but also of almost empty voids. Additionally, the underlying dark matter structure and its evolution is shown to follow the visible baryonic matter, though the luminous matter is *biased* in relation to the dark matter distribution. It is still unclear how exactly this baryonic - dark matter relation looks like, especially in the early stages of galaxy formation, and how it changed through time.

The correlation function is a commonly used tool to describe how galaxies are clustered as a function of scale, and allows to put constraints on the evolution of galaxies (Kauffmann et al. 1999, Zehavi et al. 2011). It is based on the very simple idea of measuring the probability of finding two galaxies at a given redshift (Peebles 1980). The galaxy correlation function can be interpreted by introducing two kinds of approximations. The first and most extensively used formalism is based on a simple power-law approximation of the correlation function of the form  $\xi(r) = (r/r_0)^{-\gamma}$  with two free parameters: the correlation length  $r_0$ , which describes how strongly galaxies are clustered, and the slope  $\gamma$  (Davis & Peebles 1983). The second, more recent and detailed approximation is based on halo occupation models (Seljak 2000, Peacock & Smith 2000, Magliocchetti & Porciani 2003, Zehavi et al. 2004, Zheng et al. 2005). In this framework the correlation function is built from two components, which have their influence on different scales. The *One-halo* term dominates on small scales ( $\lesssim 1.5 h^{-1}\text{Mpc}$ ) and describes clustering of galaxies which reside within dark matter haloes. The *two-halo* term describes large scale ( $\gtrsim 3 h^{-1}\text{Mpc}$ ) galaxy clustering between different haloes.

Both formulations have been extensively used in the past years. Correlation function measurements have been produced for most large galaxy surveys, such as the Sloan Digital Sky Survey (SDSS, Connolly et al. 2002, Zehavi et al. 2004), 2dF Galaxy Redshift Survey (2dFGRS, Magliocchetti & Porciani 2003), VIMOS-VLT Sky Survey (VVDS, Le Fèvre et al. 2005, Pollo et al. 2006), VIMOS Public Extragalactic Survey (VIPERS, Marulli et al. 2013), and DEEP2 (Coil et al. 2006).

Thanks to this remarkable effort it is now very well established that the strength of the galaxy clustering for the general population of galaxies is only mildly evolving from intermediate redshifts  $z \sim 1$  to  $z \sim 0$ . It has been found that galaxy clustering depends on a variety of galaxy properties like the luminosity, morphology, colour and spectral type of galaxies: luminous galaxies tend to be more clustered than faint ones and red galaxies with old stellar population are found to be more clustered than young blue ones (Norberg et al. 2002, Zehavi et al. 2004, 2012, Pollo et al. 2006, de la Torre et al. 2007, Coil et al. 2008, Quadri et al. 2008, Meneux et al. 2006, 2008, 2009, Skibba et al. 2009, Abbas et al. 2010, Zehavi et al. 2011, Coupon et al. 2012).

At redshifts higher than  $z \sim 2$  the situation is less clear. Various difficulties, mainly in collecting statistically signifi-

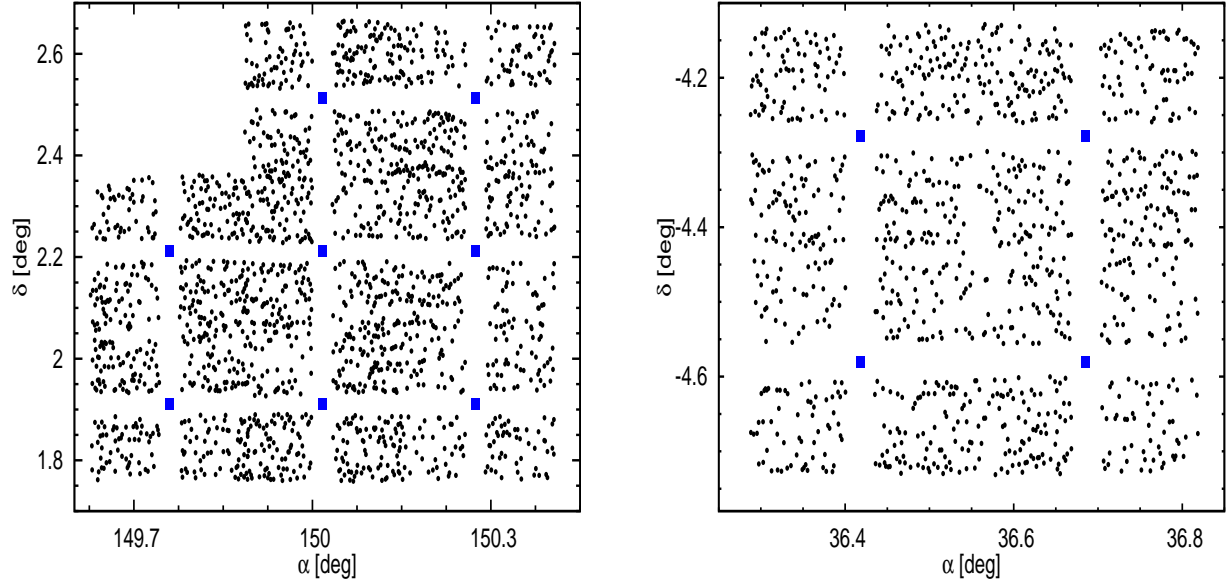
cant and representative samples, need to be overcome. While a number of attempts to measure galaxy clustering have been made at early epochs, interpreting results is not straightforward, mainly because of uncertainties connected to redshift determination, small volumes covered, and different galaxy populations selected from a range of methods. Most of the measurements at high redshifts ( $z > 1$ ) are produced using photometric surveys targeting specific classes of galaxies or applying specific observation techniques, like BzK method (Kong et al. 2006, Lin et al. 2012) or Lyman-break galaxy (LBG) selection (Foucaud et al. 2003, Ouchi et al. 2004, Adelberger et al. 2005, Kashikawa et al. 2006, Savoy et al. 2011, Bielby et al. 2013). In general, within a given sample and based on angular correlation function measurements, some evidence has been presented at  $z > 1$  showing that clustering also seems to depend on luminosity, stellar mass or colour as observed on lower redshift, with bright galaxies clustering more strongly than faint ones (Savoy et al. 2011) and passive galaxies clustering more strongly than star-forming galaxies at a given stellar mass (Lin et al. 2012). However the relation of galaxy samples used in these analyses to the general population of galaxies is not well established, which makes the study of evolution of the galaxy clustering at  $z > 2$  difficult. Moreover these galaxies can not be easily connected with the galaxy populations at lower redshifts, which make it difficult to conduct a consistent study of galaxy clustering evolution from high redshift to  $z = 0$ .

From the theoretical point of view, the correlation function should evolve with time -  $\xi = \xi(r, t)$ , because the density field of the Universe evolves over time. In the framework of the Newtonian linear perturbation theory (and in the matter-dominated era of the history of the Universe) in the co-moving coordinates the density contrast can be decomposed into time-dependent and space-dependent factors:  $\delta(\mathbf{r}, t) = D(t)\delta_0(\mathbf{r})$ , where  $\delta_0(\mathbf{r})$  is the present day value of the density contrast at a given location, and  $D(t)$  is referred to as the growth factor, which depends on the parameters of the assumed cosmological models. In practice, it means that the spatial shape of the density fluctuations in co-moving coordinates does not change, and only their amplitude increases. Consequently, the amplitude of the correlation function should increase over time:  $\xi(r, t) = D^2(t)\xi(r, t_0)$ , with a dependence on the cosmological parameters (Peebles 1980, Schneider 2006).

The consistent measurement of the evolution of a correlation function amplitude for a given galaxy population is therefore enough - in principle - to test the paradigm of the gravitational perturbations as the origin of the large scale structure of the Universe (and a way to estimate cosmological parameters). Extending these measurements to as high redshifts as possible is necessary both for the theoretical framework of the large scale structure evolution and for the galaxy formation and evolution models.

In this work we attempt to improve our current view of the evolution of the clustering of galaxies at  $z > 2$ , investigating how galaxy clustering evolved from the early phases of galaxy assembly to present times. We present a clustering analysis of the VIMOS Ultra Deep Survey (VUDS), the largest spectroscopic survey covering the redshift range  $2 < z < 6$  in a continuous way to date (Le Fèvre et al. 2014), using a sample of  $\sim 3000$  galaxies with confirmed spectroscopic redshifts in the range  $2 < z < 5$ . The sample covers a total area of  $0.81 \text{ deg}^2$  observed in two independent fields, COSMOS and VVDS-02h,

\* Based on data obtained with the European Southern Observatory Very Large Telescope, Paranal, Chile, under Large Program 185.A-0791.



**Fig. 1.** The distribution of galaxies with spectroscopic redshifts  $2 < z < 5$  in two independent VUDS fields: COSMOS on the *left* panel and VVDS-02h on *right* panel. The blue squares indicates VIMOS pointing centers.

with a mean redshift value of  $z \sim 3$ . We use a halo occupation distribution (HOD) model to put constraints on the properties of dark matter halos hosting star-forming galaxies at redshifts  $z \sim 3$ .

The paper is organized as follows. In section 2 we briefly describe the properties of the VUDS survey and our selected samples. The method used to measure the correlation function and derive power-law and HOD fits is presented in section 3. Results are described in section 4. We discuss our findings and compare to other work, both at low and high redshifts, in section 5, and we summarize our work in section 6.

We adopt a flat  $\Lambda$ CDM cosmological model with  $\Omega_m = 0.25$ ,  $\Omega_\Lambda = 0.75$  (we note that using the latest cosmology parameters from e.g. Planck Collaboration et al. 2013 does not change conclusions drawn from clustering analysis, see Zehavi et al. (2011)). The Hubble constant is normally parametrized via  $h = H_0/100$  to ease comparison with previous works, while a value  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}$  is used when computing absolute magnitudes and stellar masses. We report correlation length measurements in comoving coordinates and express magnitudes in the AB system.

## 2. The Data

### 2.1. The VUDS Survey summary

The VIMOS Ultra Deep Survey (VUDS, Le Fevre et al. 2014) is a spectroscopic survey targeting  $\sim 10,000$  galaxies performed with the VIMOS multi-object spectrograph (Le Fevre et al. 2003) at the European Southern Observatory Very Large Telescope. The main aim of the survey is to study early phases of galaxy formation and evolution at  $2 < z < 6$ . The survey covers a total area of  $1 \text{ deg}^2$  in three independent fields, reducing the effect of cosmic variance, an important property for galaxy clustering measurements. Targets are

**Table 1.** The number of galaxies in VUDS fields for each reliability flag in the redshift range  $2 < z < 5$ , as used in this study.

VUDS field	reliability flag $zflag$				Area ( $\text{deg}^2$ )
	2	3	4	9	
COSMOS	939	607	322	72	0.50
VVDS-02h	471	384	207	20	0.31
Total	1410	991	529	92	0.81

selected based on photometric redshifts ( $z_{phot} > 2.4 + 1\sigma$ ) derived from deep multi-band photometry available in the VUDS fields, augmented by colour selection (e.g. Lyman Break Galaxies, LBG) when not selected by the photometric redshift criterion, and limit of  $i_{AB} = 25$  is imposed. Details about the survey strategy, target selection, as well as data processing and redshift measurements are presented in Le Fevre et al. (2014). Here we briefly describe VUDS features which are relevant for our work.

Redshift measurements were carried out in a way similar to that developed for the VVDS survey (Le Fevre et al. 2005, Le Fevre et al. 2013), the zCOSMOS survey (Lilly et al. 2007), and VIPERS survey (Guzzo et al. 2014). The core engine for redshift measurement is cross-correlation of the observed spectrum with reference templates using the EZ redshift measurement code (Garilli et al. 2010). At the end of the process, each redshift measurement is assigned a reliability flag which expresses the reliability of the measurement:

- Flag 0: No redshift could be assigned
- Flag 1: 50 – 75% probability to be correct
- Flag 2: 75 – 85% probability to be correct
- Flag 3: 95 – 100% probability to be correct
- Flag 4: 100% probability to be correct
- Flag 9:  $\sim 80\%$  probability to be correct; spectrum has a single emission line

The VUDS sample benefits from an extended multi-wavelength dataset (see Le Fevre et al. 2014). The multi-wavelength photometry is used to compute absolute magnitudes and stellar masses from SED fitting using the ‘Le Phare’ code (Arnouts et al. 1999, Ilbert et al. 2006), as described in detail by Ilbert et al. (2013) and references therein.

## 2.2. The sample selection for clustering analysis

In this work we are using only the objects with assigned reliability flag  $zflag = 2, 3, 4, 9$  in the redshift range  $2 < z < 5$ . We use two independent fields, COSMOS and VVDS-02h, covering a total area  $0.81 \text{ deg}^2$ , as data for the third VUDS field in the ECDFS are still too sparse to be used for clustering analysis. Covered area corresponds to a volume  $\sim 3 \times 10^7 \text{ Mpc}^3$  sampled in the redshift range  $2 < z < 5$ . Our sample consists of 3022 galaxies with reliable spectroscopic redshifts in the range  $2 < z < 5$ . The sample is summarized in Table 1.

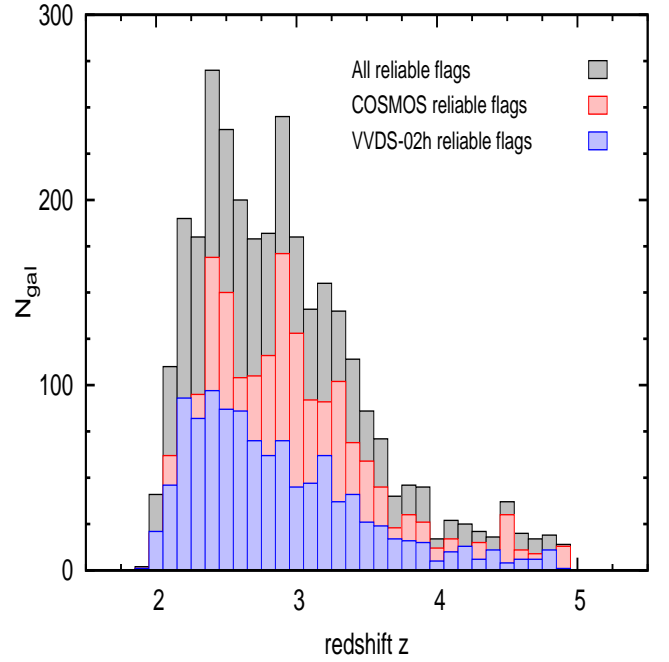
The spatial distribution of the spectroscopic galaxy sample in each of these fields is presented in Figure 1, while Figure 2 presents their redshift distributions. Mean redshift value for the whole sample is  $z \sim 2.96$ . We use 3 different redshift bins to study our sample. In a first sample we include the whole star-forming galaxy population observed by VUDS in the range  $2 < z < 5$  with a mean redshift  $z \sim 2.96$ . We also define a *low* redshift sample with a mean redshift  $\bar{z} = 2.5$  in the range  $z=[2,2.9]$  and a *high* redshift sample with  $\bar{z} = 3.45$  in  $z=[2.9,5]$ , which roughly represent two equal cosmic time bins. The general properties of these samples including median luminosity  $M_U^{\text{median}}$  and stellar mass  $\log M_*^{\text{median}}$  are listed in Table 2.

Because of the  $i_{AB} = 25$  limit of the VUDS selection, the galaxies in the high redshift sample are  $\sim 0.5 \text{ mag}$  brighter and  $\sim 0.2 M_\odot$  more massive than those in the low redshift subsample, which is the expected effect taking into account the VUDS selection strategy. The impact of these differences on the clustering measurements of our three samples are fully discussed in Section 5.

## 2.3. VUDS mock catalogues

To estimate uncertainties of the correlation function (see Sec. 3.2) and to test robustness (see App. A) of our clustering measurements we make use of a large number of mock galaxy samples, which are designed to mirror the VUDS sample in the range  $2 < z < 5$  in the most realistic way.

The 66 independent mock samples have been created based on a method which includes the halo occupation distribution (HOD) and the Stellar-to-Halo Mass Relation (SHMR). This method is the same as developed and introduced by de la Torre et al. (2013), and we refer the reader to the details presented in this paper. We provide below a brief summary of its most important features.



**Fig. 2.** The redshift distribution of the used VUDS galaxy sample for redshift range  $2.0 < z < 5.0$ . The filled gray histogram represents total sample of reliable galaxies (with  $zflag = 2 - 4, 9$ ,  $\sim 80\%$  reliability), while the red and blue histograms represent the contribution from COSMOS and VVDS2h fields respectively.

In the first step the mock catalogues containing stellar masses have been created from the MultiDark  $N$ -body simulation (Prada et al. 2012) and Pinocchio halo lightcones (Monaco et al. 2002). We followed the stellar mass to halo mass ratio (SMHR) approach based on the assumption of a monotonic relation between halo/subhalo masses and the stellar masses of the galaxies associated with them. We first populated the haloes in lightcones with subhaloes. For this we randomly distributed subhaloes around each distinct halo following a NFW profile (Navarro et al. 1997), so that their number density satisfies the subhalo mass function proposed by Giocoli et al. (2010). Then we assigned a galaxy to each halo and subhalo, with a stellar mass given by the SHMR of Moster et al. (2013).

The procedure which was followed in the next step is similar to the one used in the VVDS, zCOSMOS and VIPERS surveys also based on VIMOS observations as the VUDS survey (Meneux et al. 2006, Iovino et al. 2010, de la Torre et al. 2011b, 2013). In order to obtain fully realistic VUDS mock catalogues we add the detailed survey selection criteria (see Sec. 2.1 and Le Fevre et al. 2014). This procedure in the end produces mock parent photometric galaxy catalogues. We then apply the slit-positioning algorithm (SSPOC) with the same settings as used for the VUDS survey (Bottini et al. 2005). This allowed us to reproduce the VUDS footprint on the sky for a proper simulation of the small-scale angular incompleteness in the mocks.

As a result of this procedure, 66 realistic mock galaxy catalogues have been produced. They contain the detailed survey

completeness function and observational biases of the VUDS survey.

### 3. Clustering measurements

#### 3.1. Two-point correlation function

The two point correlation function is one of the most frequently used statistical tool used to measure and investigate galaxy clustering. It measures the probability, above Poisson, of finding two galaxies separated by a given distance  $r$  (Peebles 1980).

$$P = [1 + \xi(r)]\rho dV_1 dV_2, \quad (1)$$

where  $P$  is the probability of finding these two galaxies in two infinitely small volumes  $dV_1$  and  $dV_2$  separated by the distance  $r$ , while the average density of galaxies is given by  $\rho$ .

Although the idea is relatively simple, this definition cannot be straightforwardly applied to compute the correlation function from real data samples, because of the limitations of galaxy surveys themselves. In principle, to retrieve ideal correlation functions one would need an unlimited survey which covers the whole sky, and includes all galaxies. Naturally, in practice such surveys are unreachable. A number of estimators of  $\xi(r)$ , aimed at minimizing the effects related to the limited number of objects and limited areas covered by available surveys, have been proposed (see, e.g., Davis & Peebles 1983, Hamilton 1993). The estimator the most commonly used because of its well established capability to minimize the above mentioned observational limitations is the Landy & Szalay (1993) estimator, which we also apply in this work:

$$\xi(r) = \frac{N_R(N_R - 1)GG(r)}{N_G(N_G - 1)RR(r)} - 2 \frac{(N_R - 1)GR(r)}{N_G RR(r)} + 1 \quad (2)$$

Here,  $N_G$  and  $N_R$  represent, respectively, numbers of galaxies in the galaxy sample and randomly distributed objects in the same volume as observed in the survey;  $GG(r)$  is the number of distinct galaxy-galaxy pairs with separations lying in the interval  $(r, r + dr)$ . Similarly,  $RR(r)$  and  $GR(r)$  are the numbers of random-random pairs and galaxy-random pairs, respectively, in the same intervals. The galaxy pairs  $GG$ ,  $GR$  and  $RR$  are normalized by  $N_G(N_G - 1)$ ,  $N_G N_R$  and  $N_R(N_R - 1)$  respectively, where  $N_G$  and  $N_R$  are the number of galaxies in the data and random catalogues.

In this work we make galaxy clustering measurements using combined data from two independent VUDS survey fields - COSMOS and VVDS2h. The final  $\xi^{final}(r)$  is computed for all fields simultaneously by using Landy & Szalay estimator (Eq. 2), and the differences in size and numbers between the fields are accounted for by an appropriate weighting scheme. In particular, each pair is multiplied by the number of galaxies per unit volume for the given field:

$$\xi^{final}(r) = \frac{\sum_{i=1}^{n_{field}} w_i \cdot (GG_i(r) - 2GR_i(r) + RR_i(r))}{\sum_{i=1}^{n_{field}} w_i \cdot RR_i(r)}, \quad (3)$$

where  $w_i = (N_{gal,i}/V_i)^2$ .

The Landy & Szalay estimator requires the creation of a random catalogue, which follows the geometrical properties of the corresponding galaxy sample and is distributed in the same volume (or on the same area on the sky, in its angular version). The number of objects generated within this catalogue is also crucial - in practice it should be significantly more numerous than the real data sample. Too few random objects introduces shot noise into  $\xi(r)$  measurements, especially on small scales ( $r_p < 3h^{-1}\text{Mpc}$ ). After a set of tests, for the measurements presented in this paper we decided to use  $N_R = 100,000$ , for which additional noise was found to be negligible.

Another problem one has to face while measuring the real space correlation function is related to peculiar velocities of galaxies. In redshift space, these peculiar velocities distort distances computed from the redshifts  $z$ , and, as a result, they affect the shape of the correlation function  $\xi(r)$  itself. The corresponding distortions are known as the coherent infall and "fingers of God" (Kaiser 1987). These differences, however, apply only to the radial separations and do not have any influence on the measured position of a galaxy on the sky. A commonly applied method to eliminate them is to split the co-moving redshift space separations into two components - parallel  $\pi$  and perpendicular  $r_p$  to the line of sight, thus re-defining the redshift-space correlation function as  $\xi(r_p, \pi)$ . Integrating  $\xi(r_p, \pi)$  along the line of sight gives us a projected correlation function  $w_p(r_p)$ , which is the two-dimensional counterpart of the real-space correlation function, free from the redshift-space distortions (Davis & Peebles 1983):

$$w_p(r_p) = 2 \int_0^\infty \xi(r_p, \pi) d\pi = 2 \int_0^\infty \xi\left((r_p^2 + y^2)^{\frac{1}{2}}\right) dy. \quad (4)$$

Here,  $y$  is the real-space separation along the line of sight, and  $\xi(r)$  stands for real space correlation function computed for  $r = \sqrt{r_p^2 + y^2}$ . In practice, the upper integral limit  $\pi_{max}$  has to be finite in order to avoid adding noise to the result. After performing a number of tests, we decided to use  $\pi_{max} = 20 h^{-1}\text{Mpc}$ , as the value around which the correlation function is the most stable.

#### 3.2. Error estimates

Estimating errors of the two-point correlation function is the subject of numerous discussions in the literature since the time of the very first measurements (see Hamilton 1993, Fisher et al. 1994, Bernardeau et al. 2002). The usage of a properly constrained covariance matrix is needed to account for the fact that the values of  $w_p$  for different separations  $r_p$  are not independent - pair counts in different  $r_p$  bins include partially the same galaxies. After performing a number of tests, for the VUDS data we decided to apply a combined method, which makes use of the so-called "blockwise bootstrap re-sampling" (Barrow et al. 1984) coupled to mock catalogues (see Section 2.3), similar to the method proposed by Pollo et al. (2005).

The "blockwise bootstrap re-sampling" is based on the idea of perturbing the data set by randomly creating a large number of comparable pseudo data-sets, which differ only slightly from the original sample. A number of objects is randomly selected from the data sample (objects are allowed to be drawn multiple times), and the correlation function is computed for each of these sub-samples. This procedure is repeated  $N_{boot}$  times,

giving as a result the variation around the 'real' result, which is used as an error estimate.

The second method makes use of sets of independent mock surveys - the simulated catalogues based on large N-body simulations coupled with physical definitions of galaxies, inserted e.g. by semi-analytic models. The correlation functions computed for each independent mock catalogue give the variation of the result at each scale, and are used as error estimates.

For both the bootstrap re-sampling and mocks methods the associated covariance matrix  $\mathbf{C}$  between the values of  $w_p$  on  $i$ th and  $k$ th scale can be computed as:

$$\mathbf{C}_{ik} = \left\langle \left( w_p^j(r_i) - \langle w_p^j(r_i) \rangle_j \right) \left( w_p^j(r_k) - \langle w_p^j(r_k) \rangle_j \right) \right\rangle_j, \quad (5)$$

where " $\langle \rangle$ " indicates an average over all bootstrap or mock realizations, the  $w_p^j(r_k)$  is the value of  $w_p$  computed at  $r_p = r_i$  in the cone  $j$ , where  $1 < j < N_{\text{mock}}$  for the VUDS mocks and  $1 < j < N_{\text{boot}}$  for the bootstrap data.

In our case the covariance matrix reconstructed from  $N_{\text{mock}} = 66$  VUDS mock catalogues could not be directly applied to the observed data, because it caused the fit to be often unstable, and not able to properly converge. This instability was caused by the fact that the diagonal elements of the matrix have realistic values but the off-diagonal non-zero elements differ significantly from those pertaining to the data sample. In other words, the detailed statistical properties of the VUDS mock catalogues are not close enough to the real VUDS data to allow for such an operation directly. For this reason we computed error bars using the scatter between the VUDS mock catalogues (since we believe it is more realistic than the scatter between the bootstrap realizations), but the off-diagonal elements of the covariance matrix were computed using the classical bootstrap method. This means that for each redshift range we measured the correlation function  $w_p(r_p)$  from 1) the VUDS data; 2)  $N_{\text{mock}} = 66$  VUDS mock catalogues; 3)  $N_{\text{boot}} = 100$  bootstrap re-samplings of the VUDS data and in the fitting and error estimation we used the combination of these measurements.

### 3.3. Systematics in the CF measurement

Before introducing correlation function measurements from our data several tests have been performed on mock catalogues to investigate the influence of various VUDS survey properties on the correlation function. As the result of these tests, the correlation function correction scheme has been performed as described in the Appendix A.

In the lower panel of Figure .1 it is shown that the final measurement of the correlation function  $w_p(r_p)$ , obtained after introducing the full correction scheme, is still slightly underestimated, by as much as  $\sim 10\%$  on small scales, with respect to the 'true' correlation function obtained for the mock parent samples (see details in App. A). This systematic underestimation may affect the value of the correlation length  $r_0$ , and all HOD parameters including an underestimation of the average halo mass  $\langle M_h \rangle$  with respect to the 'true' underlying values. To estimate the size of this effect, we performed a power-law function fit on the 'true' correlation function (computed for the parent mock) as well as on the 'observed' (the mock catalogue after SSPOC selection), with the full correcting scheme, both measured as the average from 66 VUDS mock catalogues. We

found that the 'true' correlation length is on average larger than the 'observed'  $r_0$  by  $\Delta r_0 = 0.42 \pm 0.34$ , and that the 'observed' slope was underestimated by  $\Delta \gamma = 0.06 \pm 0.04$ . This, of course, would impact the galaxy bias estimation, since it is derived directly from these parameters, which we estimated to be underestimated by  $\Delta b = 0.35 \pm 0.27$ .

To address this problem we compute the ratio  $w_p^{\text{par}}/w_p^{\text{obs}}$  between the 'true'  $w_p^{\text{par}}(r_p)$  and 'observed'  $w_p^{\text{obs}}(r_p)$  correlation function (based on the VUDS mock, see appendix A) on every scale  $r_p$  for each redshift range separately. The correlation function measurements are then multiplied by this ratio for each separation  $r_p$ , resulting in the final correlation function measurement  $w_p(r_p)$  on which the power-law and HOD fitting have been performed.

All results presented in this paper include the correlation function corrections described above.

### 3.4. Power-law model

In most cases, especially in the local Universe, the correlation function  $\xi(r)$  is well described by a power law function  $\xi(r) = (r/r_0)^{-\gamma}$ , where  $r_0$  and  $\gamma$  are correlation length and slope, respectively. With this parametrisation, the integral in Equation 4 can be computed analytically and  $w_p(r_p)$  can be expressed as:

$$w_p(r_p) = r_p \left( \frac{r_0}{r_p} \right)^\gamma \frac{\Gamma(\frac{1}{2})\Gamma(\frac{\gamma-1}{2})}{\Gamma(\frac{\gamma}{2})}, \quad (6)$$

where  $\Gamma$  is the Euler's Gamma Function.

The values of  $w_p$  are not independent at different separations, hence it is not possible to use simple  $\chi^2$  minimization to find the best-fit parameters. However, the covariance matrix  $\mathbf{C}$  is symmetric and real, and therefore it can be inverted (if it is not singular). We can fit  $w_p$  by minimizing a generalized  $\chi^2$ , which is defined as

$$\chi^2 = \sum_{i,j}^{N_D} \left( w_p^{\text{obs}}(r_i) - w_p^{\text{mod}}(r_i) \right) \mathbf{C}_{ij}^{-1} \left( w_p^{\text{obs}}(r_j) - w_p^{\text{mod}}(r_j) \right). \quad (7)$$

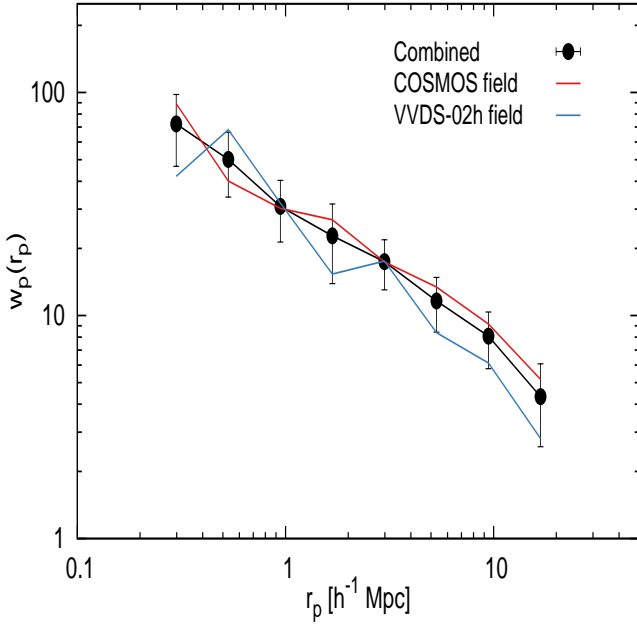
Here the covariance matrix  $\mathbf{C}$  is computed using the method described in section 3.2 (Eq. 5). The fitting procedure to estimate the power-law parameters  $r_0$  and  $\gamma$  for the projected correlation function  $w_p(r_p)$  follows Fisher et al. (1994) (see also Guzzo et al. 1997 and Pollo et al. 2005).

A more detailed description of the shape of the real-space correlation function can be done e.g. in the framework of Halo Occupation Distribution (HOD) models, which we discuss in the following section 3.5. Nevertheless the use of a power-law model to describe the correlation function remains an efficient and simple approximation of galaxy clustering properties.

### 3.5. The halo occupation distribution (HOD) model

As we mentioned above, in practice  $\xi(r)$  is usually well fitted by a power law. However, there is no strict theoretical reason to force the galaxy correlation function to assume a power law shape. Indeed, recent measurements show deviations from a power law shape. The strongest deviations are observed on small scales, especially for the most luminous galaxies (Coil et al. 2006, Pollo et al. 2006, Zehavi et al. 2011),





**Fig. 3.** The projected two-point correlation function  $w_p(r_p)$  in each individual VUDS field for the redshift range  $2 < z < 5$ . Blue and red lines correspond to the VVDS and COSMOS field, respectively. Black points and line indicate the combined correlation function from the measurements performed on galaxies from both fields simultaneously.

while probably the most famous ones are baryonic acoustic oscillations (BAOs) at large scales (Percival et al. 2010).

The small scale behaviour of the correlation function of galaxies can be well interpreted in the framework of the halo occupation distribution (HOD) models based on the relations between the distribution of dark matter and galaxies. The HOD models describe bias in terms of the probability  $P(N|M)$  that a dark matter halo of a mass  $M$  contains  $N$  galaxies of a given type. Recently this approach has been very successfully used to model the shape of the two-point correlation function (e.g. Skibba et al. 2009, Abbas et al. 2010, Zehavi et al. 2011, Coupon et al. 2012, Kim et al. 2014).

In this work we apply the halo occupation model following the current commonly used analytical prescriptions, so that our results can be easily compared to results in the literature. We describe below the most important features and components of the HOD model we use, while for the general overview of the HOD philosophy we encourage readers to look into the review paper by Cooray & Sheth (2002). A very thorough description is also given by Coupon et al. (2012).

In the HOD framework the correlation function can be split into two components. The one-halo component  $\xi^{1h}(r)$  dominates on scales smaller than the size of dark matter haloes (usually  $< 1.5 h^{-1} \text{Mpc}$ ), while the two-halo component  $\xi^{2h}(r)$  dominates on larger scales. The one-halo term arises from pairs of galaxies located within the same halo, while the two-halo terms is built by the pairs of galaxies hosted by different haloes. Consequently, the correlation function can be written as:

$$\xi(r) = \xi^{1h}(r) + \xi^{2h}(r). \quad (8)$$

The properties of the first term, i.e.  $\xi^{1h}(r)$  can be described with the use of the halo occupation model which tries to follow the distribution of galaxies located in one dark matter halo. The second term  $\xi^{2h}(r)$ , depends on the statistical properties of the large scale dark matter density field and the distribution of dark matter haloes with respect to this density field.

The most important properties of the dark matter halo from this point of view are the halo mass function  $n(M, z)$ , for which we adopted the fitting formula proposed by Tinker et al. (2008), the halo density profile  $\rho(r|M)$ , for which we assumed the form described by Navarro, Frenk, & White (1997), and the halo bias  $b_h(M, z, r)$ , for which we use the formula proposed by Tinker et al. (2010) with the scale-dependency from Tinker et al. (2005).

We parametrized our halo occupation model in the way used by Zehavi et al. (2005b) and motivated by Kravtsov et al. (2004). In particular, we express the halo occupation function, i.e. the number of galaxies per halo, as a sum of a central galaxy and satellite galaxies. The mean occupation function for central galaxies  $N_c(M)$  is represented by a step function, while the mean halo occupation function for satellite galaxies  $N_s(M)$  is approximated by a Poisson distribution with the mean being a power-law:

$$\begin{aligned} \langle N_g|M \rangle &= 1 + \left( \frac{M}{M_1} \right)^\alpha \quad \text{for } M > M_{min} \\ &= 0 \quad \text{otherwise,} \end{aligned} \quad (9)$$

where  $M_{min}$  is the minimum mass needed for a halo to host one central galaxy, and  $M_1$  is the mass of a halo having on average one satellite galaxy, while  $\alpha$  is the power law slope of the satellite mean occupation function.

The HOD model used in this work is simplified in comparison with the ones widely used at lower redshift ranges, e.g. by Zehavi et al. (2011) for fitting the SDSS data. However, based on the available statistics of the VUDS sample, with lower numbers of galaxies than e.g. in the SDSS, the three parameter model appears to be the best solution to retrieve robust measurements of  $M_{min}$  and  $M_1$ . A similar approach was taken by Abbas et al. (2010) in the case of the measurements from the VVDS survey at  $z \sim 1$  with a similar sample size (Abbas et al. 2010). Since the details of the interplay between galaxy evolution and large scale structure growth at epochs as early as the ones probed by VUDS are still poorly understood, it is reasonable to keep a simplified description and not to dilute our analysis with models introducing too many free variables. The main results presented in this paper, however, are not expected to differ much from results where more complex HOD models are applied.

From the best-fit HOD parameters it is possible to obtain quantities describing halo and galaxy properties, like the average host halo mass  $M_h$

$$\langle M_h|g \rangle(z) = \int dM M n(M, z) \frac{\langle N_g|M \rangle}{n_g(z)}, \quad (10)$$

the large-scale galaxy bias  $b_L^{HOD}$

$$b_L^{HOD}(z) = \int dM b_h(M) n(M, z) \frac{\langle N_g|M \rangle}{n_g(z)} \quad (11)$$

where  $n(M, z)$  is the dark matter mass function,  $b_h(M, z)$  is the large-scale halo bias, and the  $n_g$  given by:

$$n_g = \int n(M) \langle N_g | M \rangle dM \quad (12)$$

represents the number density of galaxies.

### 3.6. Fitting the HOD model to the correlation function

To explore the HOD parameter space we implemented our HOD model with the publicly available CosmoPMC<sup>1</sup> code, which is using the "Population Monte Carlo" (PMC) technique to sample likelihood space (Wraith et al. 2009, Kilbinger et al. 2011). PMC is the adaptive importance-sampling technique (Cappé & Moulines 2007) which allows efficient sampling of the parameters space for the large number of samples. For each galaxy sample we fit the projected correlation function  $w_p(r_p)$  and the number density of galaxies  $n_g$ , by summing both contributions in log-likelihood obtained by:

$$\chi^2 = \sum_{i,j}^{N_D} (w_p^{obs}(r_i) - w_p^{mod}(r_i)) C_{ij}^{-1} (w_p^{obs}(r_j) - w_p^{mod}(r_j)) + \frac{(n_g^{obs} - n_g^{mod})^2}{\sigma_{n_g}^2} \quad (13)$$

where  $n_g^{mod}$  is given by Eq. 12 at the mean redshift of the sample. The data covariance matrix is approximated taking Eq. 5. The error on the galaxy number density  $\sigma_{n_g}$  contains Poisson noise and cosmic variance.

### 3.7. Galaxy large-scale bias

According to the current cosmological paradigm of structure formation, galaxies form and evolve inside dark matter halos (White & Rees 1978). In other words, there exists a connection between the dark matter distribution and galaxies in the dense dark matter regions, where the halo clustering is important, galaxies should be more clustered. The galaxy spatial distribution, however, is biased with respect to the dark matter density field. The strength of this effect is referred to as *galaxy bias*.

As a first approximation we use a linear model for the galaxy large-scale bias. It assumes a linear relation between galaxy  $\sigma_{R,g}(z)$  and mass  $\sigma_{R,m}$  rms at a given redshift:

$$\sigma_{R,g}(z) = b \sigma_{R,m}(z) \quad (14)$$

where  $b$  is the galaxy bias. We also assume that  $b$  is independent of scale  $R$ , which is true especially for large scales. Usually the adopted value for the scale on which  $\sigma$  is measured is  $R = 8h^{-1}\text{Mpc}$ . Locally, at  $z = 0$ , the mass fluctuations reach the value  $\sigma_{8,m}(z = 0) = 0.83$  (Planck Collaboration et al. 2013). In our model the redshift evolution of this quantity is described as:

$$\sigma_{8,m}(z) = \sigma_{8,m}(z = 0) D(z), \quad (15)$$

where

$$D(z) = \frac{g(z)}{g(0)(1+z)}, \quad (16)$$

and  $g(z)$  is the normalized *growth factor*, which describes how fast the linear perturbations grow with the scale factor. Its value depends on the assumed cosmological model, i.e. cosmological parameters. After Carroll et al. (1992) we write down

$$g(z) = 2.5 \Omega_m \left[ \Omega_m^{4/7} - \Omega_\Lambda + \left( 1 + \frac{\Omega_m}{2} \right) \left( 1 + \frac{\Omega_\Lambda}{70} \right) \right]^{-1}, \quad (17)$$

where  $\Omega_m$  and  $\Omega_\Lambda$  are the matter and dark energy density parameters at a given redshift.

The galaxy rms at a given scale  $R$  can be retrieved from the power-law fits to the correlation function as,

$$\sigma_{8,g} = \sqrt{C_\gamma \left( \frac{r_0}{8h^{-1}\text{Mpc}} \right)^\gamma}, \quad (18)$$

with

$$C_\gamma = \frac{72}{2\gamma(3-\gamma)(4-\gamma)(6-\gamma)}. \quad (19)$$

where  $r_0$  and  $\gamma$  are the correlation length and the slope of power-law approximation of the correlation function.

In the following sections, we refer to the galaxy large-scale bias computed by this method as  $b_L^{PL}$ . This is also used to compute the values based on measurements of the correlation function from other surveys when the  $b_L$  estimations were not presented by the authors; in these cases we use their power-law best fit parameters  $r_0$  and  $\gamma$  (e.g. from Le Fèvre et al. 2005). Another estimate of linear large-scale galaxy bias can be obtained from the HOD model. We compute the large-scale galaxy bias  $b_L^{HOD}$  using the fit of the halo occupation function (Eq. 11). The comparison of these two bias values and differences between these two procedures are discussed in section 5.

## 4. Results

### 4.1. The VUDS correlation function - general population

The projected two-point correlation function  $w_p(r_p)$  is computed for the galaxy population with  $i_{AB} < 25$  at  $2 < z < 5$  separated into three redshift ranges. We are using a total number of 3 022 spectroscopically confirmed VUDS galaxies located in two fields. In Figure 3 we present composite correlation function  $w_p(r_p)$  measurements for the full data sample in the redshift range  $2 < z < 5$ , along with the measurements made for each field independently. It is worth stressing that it is the first time that the correlation function has been computed with such a high accuracy from spectroscopic data at redshift  $z \sim 3$ . As expected, the values of the correlation function measured for galaxies located in separate fields are slightly different due to cosmic variance. However, for both fields the measurements are consistent with respect to the error bars and the signal is retrieved on all scales  $0.3 < r_p < 16 h^{-1}\text{Mpc}$ , so that the composite function is not dominated by the signal from any of the individual fields at any separation  $r_p$ .

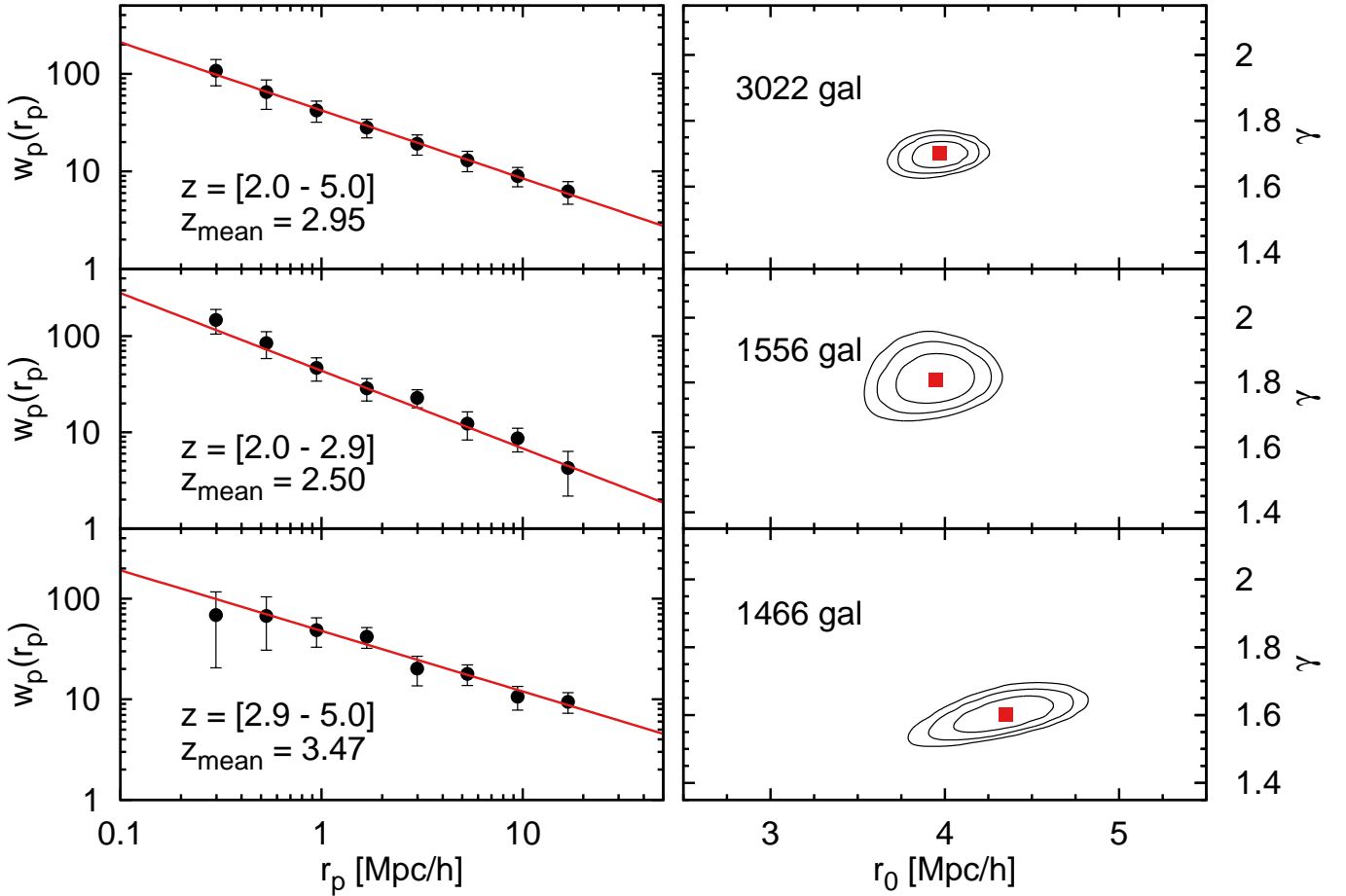
The correlation functions for the full sample  $z=[2,5]$  and the two  $z=[2,2.9]$  and  $z=[2.9,5]$  redshift intervals are presented in Figure 4.

<sup>1</sup> <http://www2.iap.fr/users/kilbinge/CosmoPMC/>



**Table 2.** Properties of general population VUDS samples in the redshift range  $2.0 < z < 5.0$ .

$z$ range	$z_{\text{mean}}$	$M_U^{\text{median}}$	$\log M_*^{\text{median}}$	$N_{\text{gal}}$	$r_0$	$\gamma$	$b_L^{\text{PL}}$
[2.0 – 5.0]	2.95	-21.55	9.76	3022	$3.97^{+0.36}_{-0.38}$	$1.70^{+0.09}_{-0.09}$	$2.68^{+0.22}_{-0.21}$
[2.0 – 2.9]	2.50	-21.31	9.66	1556	$3.95^{+0.48}_{-0.54}$	$1.81^{+0.02}_{-0.06}$	$2.39^{+0.32}_{-0.29}$
[2.9 – 5.0]	3.47	-21.81	9.86	1466	$4.35^{+0.60}_{-0.76}$	$1.60^{+0.12}_{-0.13}$	$3.26^{+0.47}_{-0.37}$

**Fig. 4.** *Left panel:* The projected two-point correlation function  $w_p(r_p)$ . The symbols and errorbars denote measurements of the composite correlation function from the VUDS survey. *Right panel:* The associated best-fit power-law parameters  $r_0$  and  $\gamma$  along with error contours for the general galaxy population separated in three redshift ranges.

#### 4.2. General population - Power law modelling

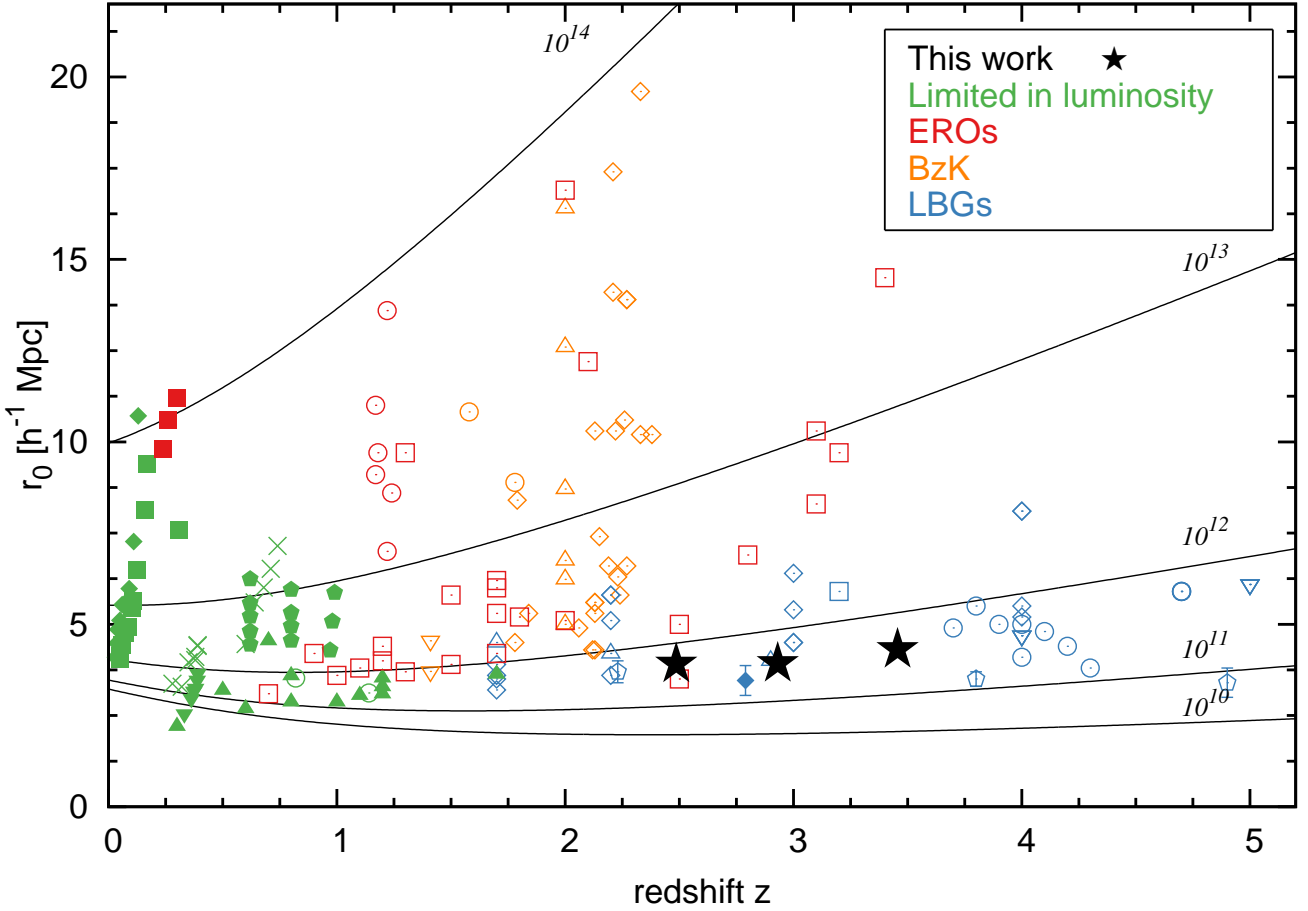
We fit a power-law to  $w_p(r_p)$ , with two free parameters  $r_0$  and  $\gamma$ , to quantify the clustering strength. All the points in the range  $0.3 \, h^{-1} \text{Mpc} < r_p < 16 \, h^{-1} \text{Mpc}$  are used (see Sec. 3.4 for the fitting method details). The measured best-fit parameters are listed in Table 2 and the power-law fit of the correlation function for each redshift sub-sample is shown in Figure 4 together with the  $(r_0, \gamma)$  error contours.

Within our sample we find  $r_0 \sim 3.95^{+0.48}_{-0.54} \, h^{-1} \text{Mpc}$  in  $z=[2.0, 2.9]$  and  $r_0 \sim 4.35^{+0.60}_{-0.76} \, h^{-1} \text{Mpc}$  in  $z=[2.9, 5]$ . The slight increase of  $r_0$  with redshift is probably due to the small

luminosity differences between samples (see Sec. 5 for discussion), although the difference is marginally significant given measurement errors. For these samples, the slope  $\gamma$  varies between  $\gamma = 1.60^{+0.12}_{-0.13}$  and  $\gamma = 1.81^{+0.02}_{-0.06}$ , hence showing a tendency to decrease with redshift.

#### 4.3. General population - HOD modelling

The best-fit HOD model defined in section 3.5 is fit to each of three redshift sub-samples as presented in the left panels of Figure 6. The HOD model provides a good fit to the real-space correlation function  $w_p(r_p)$  in all the redshift ranges, particu-



**Fig. 5.** Correlation length  $r_0$  as a function of redshift from various surveys of different types of objects. Black stars indicate the measurements obtained from this work in three redshift ranges. Different colours indicate different galaxy populations targeted by using various techniques, as indicated in the upper right corner. Open symbols indicate measurements based on photometric data, while filled symbols are for measurements from spectroscopic data. *Blue*: LBG galaxies (Foucaud et al. 2003, Ouchi et al. 2004, Adelberger et al. 2005, Kashikawa et al. 2006, Savoy et al. 2011, Bielby et al. 2013, Barone-Nugent et al. 2014). *Orange*: BzK galaxies (Blanc et al. 2008, Hartley et al. 2010, McCracken et al. 2010, Lin et al. 2012). *Green*: Galaxy samples from surveys limited in luminosity (Norberg et al. 2002, Coil et al. 2006, Le Fèvre et al. 2005, Pollo et al. 2006, Zehavi et al. 2011, Marulli et al. 2013, Skibba et al. 2009). *Red*: EROs or massive red galaxies (Daddi et al. 2003, Zehavi et al. 2005a, Brown et al. 2008). The five solid curves show the correlation length of dark halos with different minimum masses, as labelled.

larly in  $z=[2,2.9]$  where it is an excellent representation of the data. For the highest redshift bin  $z=[2.9,5]$  the fit is not as good. This is mainly due to the more noisy measurement of the correlation function in this redshift range.

The three HOD parameters, the minimum halo mass  $M_{min}$ , the satellite occupation halo mass  $M_1$ , and the high halo mass slope  $\alpha$  (see sec. 3.5), are inferred from the full covariance matrix (see section 3.6) and are listed in Table 3. The parameters are allowed to vary within a large range set from previous observations at high and low redshift:  $10 < \log M_{min} < 14$ ,  $10 < \log M_1 < 15$ ,  $0.6 < \alpha < 2.0$ . The right panel of Figure 6 presents the halo occupation function obtained from the HOD fit within three redshift ranges. We do not observe any significant difference in halo masses between the two redshift ranges probed for the general VUDS galaxy population. The minimum halo mass, for which a halo hosts at least one central galaxy

$M_{min}$ , has comparable value for the  $z \sim 2.5$  and  $z \sim 3.5$  samples.

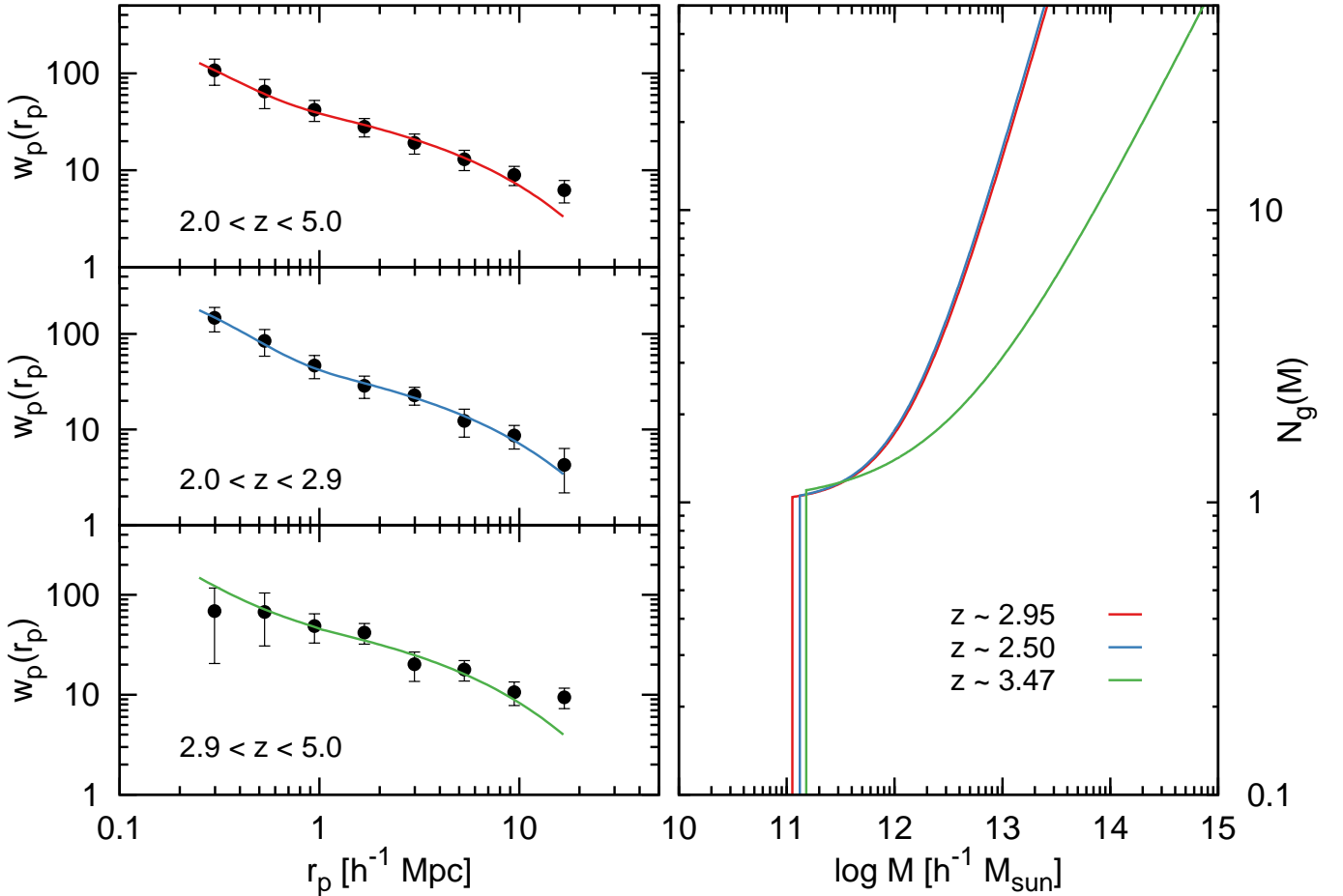
The satellite occupation mass  $M_1$  is noticeably higher in the higher redshift bin compared to the lower redshift sample, although the errors are quite large. These uncertainties are related to the weak one-halo term signal in the correlation function for the higher redshift measurement.

The high halo mass slope  $\alpha$  of the satellite mean occupation function is around  $\sim 1.3$  for  $z=[2,2.9]$ , significantly steeper than  $\alpha = 0.73$  found in  $z=[2.9,5]$ .

The average host halo mass  $\langle M_h | g \rangle$ , large-scale galaxy bias  $b_L^{HOD}$  and associated  $1\sigma$  errors, are reported in Table 3. The redshift evolution of these parameters is presented in Figure 7 - for the average host halo mass, and in Figure 8 - for the large-scale galaxy bias. These results are fully discussed in Section 5.

**Table 3.** The HOD and and HOD-based parameters for the redshift sub-samples

$z$ range	$\log M_{\min}$	$\log M_1$	$\alpha$	$\log \langle M_h \rangle$	$b_L^{HOD}$
[2.0 – 5.0]	$11.04^{+0.33}_{-0.36}$	$12.11^{+0.51}_{-0.45}$	$1.29^{+0.12}_{-0.11}$	$11.75^{+0.23}_{-0.28}$	$2.82^{+0.27}_{-0.16}$
[2.0 – 2.9]	$11.12^{+0.23}_{-0.35}$	$12.09^{+0.32}_{-0.35}$	$1.30^{+0.08}_{-0.10}$	$12.01^{+0.16}_{-0.18}$	$2.55^{+0.16}_{-0.21}$
[2.9 – 5.0]	$11.18^{+0.56}_{-0.70}$	$12.55^{+0.85}_{-0.88}$	$0.73^{+0.23}_{-0.30}$	$11.61^{+0.46}_{-0.40}$	$3.48^{+0.31}_{-0.26}$

**Fig. 6.** *Left panel:* The correlation function for three redshift sub-samples as in Figure 4. The solid lines present the best-fit halo occupation model. *Right panel:* The evolution of the halo occupation for the three redshift ranges.

#### 4.4. Linear large-scale bias measurements

The large-scale linear galaxy bias  $b_L^{PL}$  has been computed using the method described in Sec. 3.7, with the best-fit correlation parameters obtained for each galaxy sub-sample. The bias values obtained and the associated  $1\sigma$  errors are listed in Table 2. We find that the galaxy bias is increasing from  $2.39 \pm 0.32$  at  $z \sim 2.5$  to  $3.26 \pm 0.47$  at  $z \sim 3.5$ , galaxies in the higher redshift sample appearing to be more biased than those in the lower redshift sample. Additionally, we note that the bias reaches a relatively high value at  $z \sim 3.5$ . This is further discussed in Sec. 5.

## 5. Discussion

### 5.1. Evolution of the clustering length $r_0$ from $z \sim 3.5$ until the present epoch

In this paragraph we discuss the evolution of clustering properties of the general galaxy population since a redshift  $z \sim 3.5$  (see Sec. 4) in the view of other measurements presented in the literature.

Within our sample, we observe the clustering strength to increase slightly with redshift over the range  $z = [2.0 - 5.0]$ , with  $r_0$  being the highest at the highest redshift  $z \sim 3.5$ . However, this increase is marginally significant given the measurement

errors. The observed increase of  $r_0$  with  $z$ , surprising at the first glance, may be caused by the luminosity difference between two samples at  $z \sim 2.5$  and  $z \sim 3.5$ . Due to the magnitude selection  $i_{AB} \leq 25$  of the VUDS sample we observe systematically more luminous galaxies at higher  $z$ . As a result, our high redshift sample contains on average more luminous galaxies, with median  $M_U = -21.81$ , while for the lower redshift sample the median  $M_U = -21.31$ . The luminosity dependence of galaxy clustering is a well established fact; it has been observed both at early and late cosmic epochs (e.g. Zehavi et al. 2011, Coupon et al. 2012, Abbas et al. 2010, Pollo et al. 2006) and luminous galaxies at all  $z$  investigated so far were found to be more strongly clustered than the fainter ones which is usually related to the fact that they are located in more massive dark matter haloes and, consequently, stronger fluctuations of the dark matter density field. If this trend is also true for the redshift range investigated in this work, the observed increase of  $r_0$  with  $z$  may be caused by the fact that our higher redshift sample is on average more luminous than the lower one. The dependency of clustering on luminosity at these high redshifts will be investigated in a future paper (Durkalec et al. in prep).

The correlation power-law parameters  $r_0 = 3.97 \pm 0.48$  and  $\gamma = 1.70 \pm 0.09$  at  $z \sim 3$  are comparable with the clustering parameters for the star forming galaxies at intermediate ( $z \sim 1$ ) redshift (e.g. Le Fèvre et al. 2005, Pollo et al. 2006), but the clustering length is lower than found for galaxies in the local Universe (e.g. Zehavi et al. 2011, Norberg et al. 2002). This has to be placed in the global perspective of the clustering expected to grow stronger along cosmic time.

A direct comparison of clustering strength at different epochs is a difficult task for several reasons. Different measurements are based on various galaxy samples, each with a unique selection function. Since galaxies cluster differently depending on their properties (like luminosity, colours or stellar masses), and in practice every catalogue contains a galaxy population somewhat different, the differences in clustering measurements may primarily be attributed to different galaxy sample properties as much as to evolutionary effects. The inhomogeneities between galaxy samples are especially pronounced for the high redshift samples. At  $z > 2$  correlation functions have mostly been measured from surveys targeting specific classes of galaxies, e.g. extremely massive red objects or sources selected using a Lyman-break technique selecting star-forming galaxies. In addition, until recently most clustering measurements at  $z > 2$  were produced from the angular correlation function computed on photometric samples with the knowledge of the redshift being only approximate and based on color selection (BzK e.g. Daddi et al. 2003, McCracken et al. 2010; LBG e.g. Adelberger et al. 2005; LAE e.g. Ouchi et al. 2004). Compared to spectroscopic samples, the lower accuracy and precision in redshift measurements in these samples and contamination in a given redshift bin from galaxies at other redshifts introduces a blurring of large scale structures which is difficult to overcome even when using large samples.

Our  $r_0$  estimations are in excellent agreement with other available results at similar high redshifts (e.g. Daddi et al. 2003, Ouchi et al. 2004, Adelberger et al. 2005, Kashikawa et al. 2006, Hartley et al. 2010, Savoy et al. 2011), as presented in Figure 5. While  $r_0$  in the VUDS sample is on the lower side of previous clustering measurements at these redshifts, most other samples have specific selection functions

from which it is expected that they have a higher clustering bias than our sample, particularly for the redder or more luminous samples. Our measurements can be compared to the measurements by Bielby et al. (2013), based on a large sample of 2135 spectroscopically confirmed LBGs in  $\sim 10\,000$  arcmin<sup>2</sup> at  $z = 2.8$ , the largest sample with a spectroscopically-based measurement from a large volume at these redshift prior to our measurements. They find a clustering length  $r_0 = 3.46 \pm 0.41$  at  $z = 2.79$  and a slope  $\gamma = 1.5 - 1.6$ , flatter than for local galaxies, in good agreement with our measurements given the associated errors. Savoy et al. (2011) measured a correlation length  $r_0 = 3.5 \pm 0.5$  from  $\sim 1400$  UV-selected galaxies with photometric redshifts  $z \sim 3$  and  $23 < R < 27$  in the Keck Deep Fields, and Geach et al. (2012) found  $r_0 = 3.7 \pm 0.3$  from the clustering of H $\alpha$  emitters (HAEs) at  $z = 2.23$  from HiZELS, both results being very similar to ours at comparable redshifts. We conclude that our measurements are in general agreement with the most recent and least biased measurements available in the literature. Our results are among the most reliable today given the VUDS survey area and the number of galaxies used in our study.

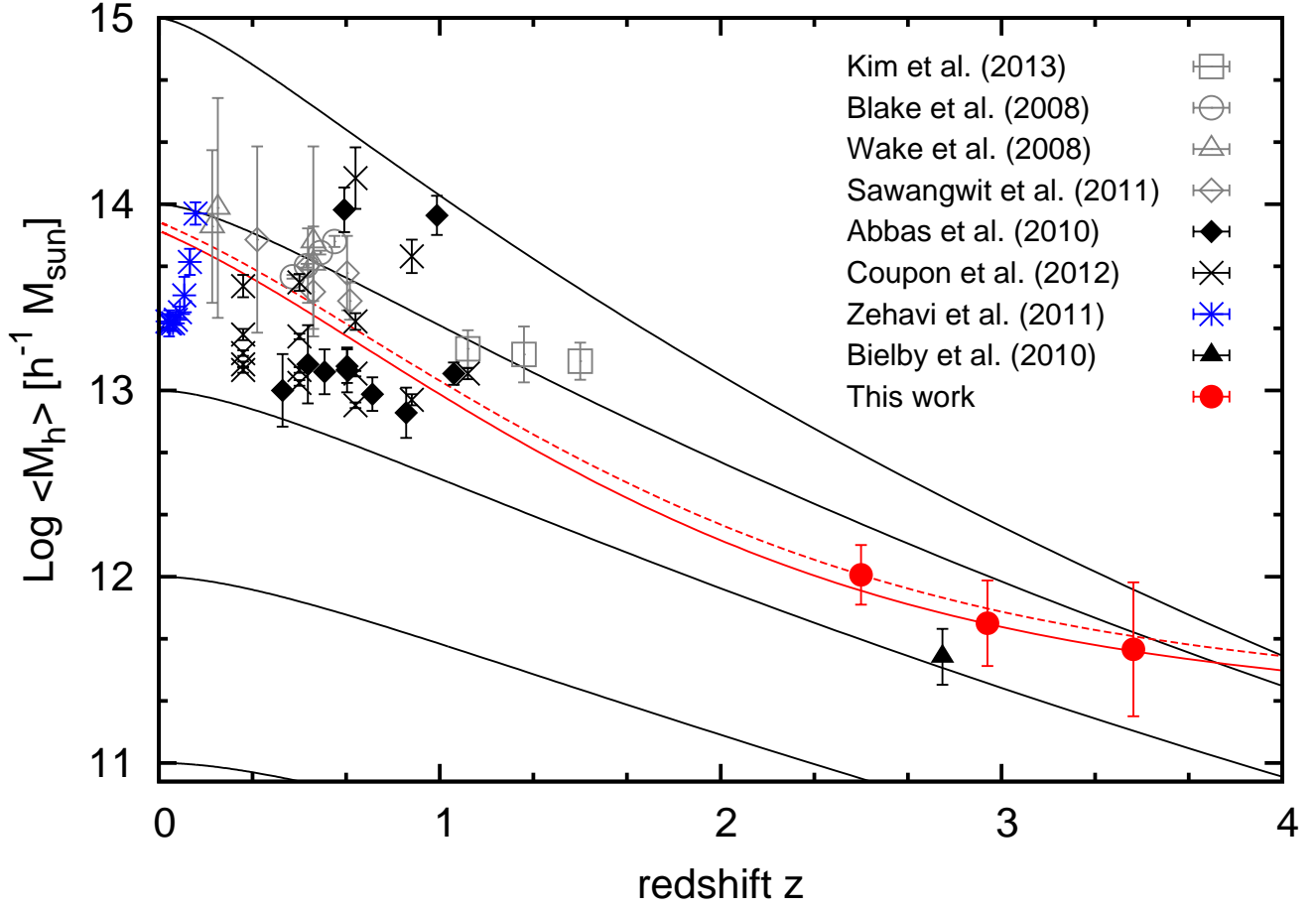
## 5.2. The host halo mass

We draw in Figure 5 the model predictions for the evolution of  $r_0$  of dark matter halos following the prescription given by Mo & White (2002) (which is based on the Press & Schechter (1974) formalism). According to this model, dark matter halos of an average mass  $\langle M_h \rangle \sim 10^{11.5} M_\odot$  have the same  $r_0$  as observed VUDS galaxies at  $z \sim 3$ . If we take the rough assumption that at  $z \sim 3$  haloes are occupied mostly by only one galaxy (central), this mass would be the average mass of a halo hosting the typical VUDS galaxy. However, this result is only a first approximation, since it is based on simplified assumptions of the Press & Schechter (1974) theory and needs to be refined.

The halo occupation model (HOD) is expected to give more accurate predictions, since it accounts for the actual number and distribution of galaxies occupying haloes. We obtain the number weighted average mass  $\langle M_h \rangle$  of haloes hosting VUDS galaxies using the best-fit HOD parameters (Eq. 10), in three redshift ranges. We find that the halo mass decreases with redshift (see Figure 7), evolving from  $\log \langle M_h \rangle = 11.61^{+0.46}_{-0.40} h^{-1} M_\odot$  at  $z \sim 3.5$  to  $\log \langle M_h \rangle = 12.01^{+0.16}_{-0.18} h^{-1} M_\odot$  at  $z \sim 2.5$ . These estimations are in reasonable agreement with the recent observations by Bielby et al. (2013) who found an average mass of  $\log \langle M_h \rangle \sim 11.57 \pm 0.15 M_\odot$  for haloes hosting LBG galaxies at a redshift  $z \sim 2.79$ .

It can also be seen from Fig. 7 that our estimated host halo masses are much lower than the ones typically observed in the local Universe. At  $z \sim 3$  the typical host halo masses are located in the range between  $\langle M_h \rangle \sim 10^{11} - 10^{12} M_\odot$  while locally these masses usually reach the values between  $10^{13} - 10^{14} M_\odot$ . These results indicate that at some stage dark matter halos must have experienced a rapid accretion phase, as expected in the framework of the hierarchical mass assembly.

A universal formula for the mass growth of cold dark matter halos has been derived by van den Bosch (2002). The history of



**Fig. 7.** The evolution of the number-weighted average host halo mass given by Eq. 10 for the three redshift ranges analysed in this study. The red filled circles indicate mass estimations from VUDS. Black and grey symbols represent the results of previous work based on spectroscopic and photometric surveys respectively. The solid black lines indicate how a host halo of a given mass  $M_0$  at  $z = 0$  evolves with redshift, according to the model given by van den Bosch (2002). The solid red line represents the halo mass evolution derived using Eq. 21, with the HOD parameters obtained from the best-fit HOD model at a redshift  $z \sim 3$ . The dashed red line is using the HOD best-fit parameters for  $z \sim 2.5$ . VUDS galaxies with a typical  $L_\star$  luminosity are likely to evolve into galaxies with a luminosity  $> L_\star$  today.

a halo with a given mass  $M_0$  at  $z = 0$  can be traced back in time using a simple formula:

$$\log\langle\Psi(M_0, z)\rangle = -0.301 \left[ \frac{\log(1+z)}{\log(1+z_f)} \right]^\nu, \quad (20)$$

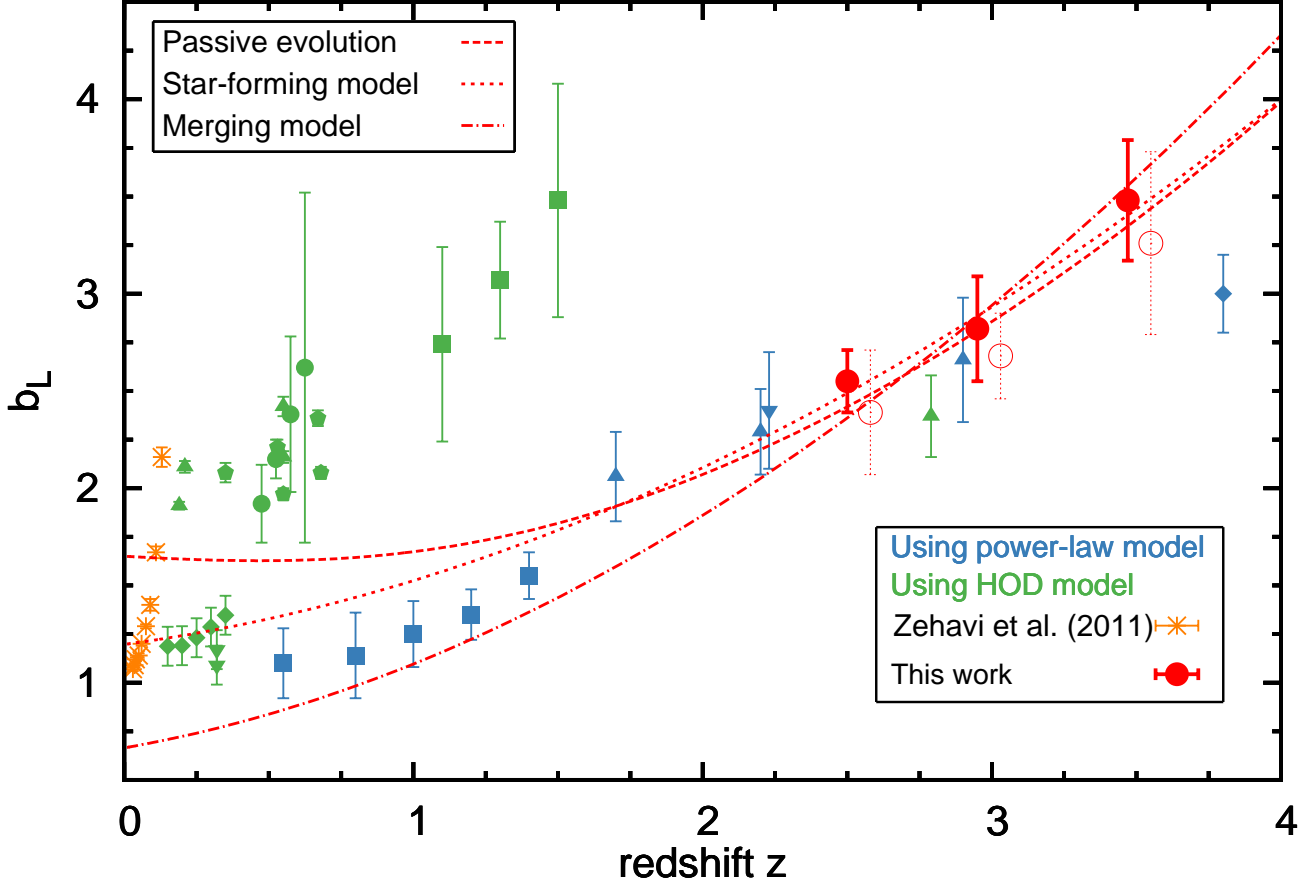
where  $z_f$  and  $\nu$  are free fitting parameters that depend on halo mass and cosmological parameters. To obtain these fitting parameters we follow the analytical formula from van den Bosch (2002). We then trace the evolution of the galaxy population sampled by VUDS from  $z \sim 3$  to the present epoch to predict the mass of halos hosting the present day descendants of the VUDS galaxies. We find that in this model the typical VUDS halo with a mass  $\langle M_H \rangle \sim 10^{11.75} M_\odot$  at  $z \sim 3$  should evolve into a halo with a mass  $\langle M_H \rangle \sim 10^{13.5} M_\odot$  at  $z = 0$ . In the local SDSS galaxy sample Zehavi et al. (2011) found that halos with these masses are typically occupied by star forming galaxies with luminosity  $M_r < -20.5$ . According to the van den Bosch (2002) model the star forming galaxies at  $z \sim 3$  in VUDS with

a typical characteristic luminosity  $L_\star$  would likely evolve into galaxies equivalent to or brighter than  $L_\star$  at the present day.

The above comparison assumes that each halo is occupied by only one galaxy. This is not expected to generally be the case, and the above picture, while broadly correct might need to be adjusted. In order to trace the evolution of dark matter haloes and the hosted galaxy population in a more realistic way we use both the halo mass growth model  $\Psi(M_0, z)$  and the halo occupation function  $\langle N_g | M \rangle$  at redshift  $z = 3$ . The average halo mass  $M_h$  as a function of redshift  $z$  is measured taking (see Eq. 10 for comparison):

$$\langle M_h \rangle(z) = \int dM \Psi^{-1}(M, z) n(M, z) \frac{\langle N_g | M \rangle}{n_g(z)}, \quad (21)$$

here the  $\Psi^{-1}(M, z)$  is the inverse mass growth function proposed by van den Bosch (2002) (Eq. 20),  $n(M, z)$  is the dark matter mass function, and  $n_g$  the galaxy number density. This allows us to trace the history of the typical dark matter halo hosting the average VUDS galaxy from a redshift  $z = 3$  to the



**Fig. 8.** Large-scale linear galaxy bias retrieved by using two methods. Red filled circles indicate measurements of  $b_L^{HOD}$ , while the open red circles represent  $b_L^{PL}$  in three redshift ranges. Green and blue point represents the results from previous works, retrieved by using HOD (Blake et al. 2008, Wake et al. 2008, Ross & Brunner 2009, Sawangwit et al. 2011, Kim et al. 2014) and power-law frameworks (Marinoni et al. 2005, Adelberger et al. 2005, Geach et al. 2012, Barone-Nugent et al. 2014, Bielby et al. 2013). We show 3 different models for the evolution of bias: a *passive* evolution model (red dashed line) a *merging* model (red dot-dashed line), a *star-forming* model (dotted line), as described in the Sec. 5.3.

present day, as presented in Figure 7. We find that the typical halo of mass  $\langle M_h \rangle = 10^{11.75}$  at  $z \sim 3$  should evolve into a halo of mass  $\langle M_h \rangle = 10^{13.9}$  at redshift  $z = 0$ . Comparing again our results from this improved model with Zehavi et al. (2011), halos with these masses are typically occupied by star forming galaxies with  $M_r < -21.5$ , above the characteristic luminosity  $M_\star$  at  $z \sim 0$ . This means that the general galaxy population observed at redshift  $z = 3$  evolves into the brightest and most massive galaxies, which occupy the most massive halos observed in the local Universe, as is expected in the hierarchical mass growth paradigm.

### 5.3. Galaxy bias

The hierarchical model of galaxy formation predicts an increase of the galaxy bias with redshift Mo & White (1996). In this section we discuss the relation between the distribution of galaxies and the underlying dark matter density field at different redshifts, by comparing the galaxy bias values measured

from VUDS to the bias of galaxy populations at different redshifts from the literature.

Figure 8 shows the linear large-scale galaxy bias  $b_L$ , computed using methods based on power-law  $b_L^{PL}$  (see Sec. 3.7) and HOD  $b_L^{HOD}$  approximations (see Sec. 3.5). The results obtained from both methods are comparable within errors, and small differences are likely due to the less accurate power-law approximation. The general VUDS galaxy population shows a bias decreasing with decreasing redshift. According to our measurements, from  $z \sim 3.5$  to  $z \sim 2.5$  the linear bias decreased by  $\Delta b_L \sim 0.9$ . Both our values at  $z = 2.5$  and  $z = 3.5$  are significantly higher than measurements at later epochs reported in the literature. For comparison, we plot the linear large-scale bias inferred for the intermediate ( $z = [0.5 - 1.5]$ ) redshift range by Marinoni et al. (2005), based on VVDS-Deep survey, for  $z > 2$  by Geach et al. (2012) based on observations of  $H\alpha$  emitters, and by Barone-Nugent et al. (2014) for  $z \sim 3.8$  LBGs, along with the values we computed using the best-fit power-law parameters  $r_0$  and  $\gamma$  from Le Fèvre et al. (2005), Adelberger et al. (2005) and Bielby et al. (2013) (the method is described in sec-



tion 3.7). The values presented in Table 3 are similar to the bias measurements recently observed for *LBGs* at redshift  $z > 2$ . Bielby et al. (2013) found  $b_{gal} = 2.33 \pm 0.17$  from their LBG-selected spectroscopic sample, and at an even higher redshift  $z \sim 3.8$  Barone-Nugent et al. (2014) estimated  $b_L = 3.0 \pm 0.2$  for similarly selected galaxies.

The visible decrease of the galaxy bias with cosmic time can be explained in terms of the hierarchical scenario of structure formation. In this framework galaxies formed at early epochs in the most dense regions (highest peaks of the density field) which are the most biased with respect to the underlying average mass density field. As the mass density field evolved with time, galaxy formation is then expected to systematically move into less dense, hence less biased, regions. Additionally, it is likely that the gas in the most dense regions became too hot for collapse, and thus star formation could not take place in the strongest over-densities any more (e.g. Blanton et al. 1999). Therefore, the observed decrease of galaxy bias with cosmic time may be explained as galaxy formation is moving to less dense areas as cosmic time increases in a downsizing trend (De Lucia et al. 2006).

In Figure 8 we also draw the predicted evolution of the large-scale galaxy bias using different prescriptions. We consider three different theoretical descriptions of the biasing functions, based on different ideas of how evolution may proceed: the *passive*, the *merging*, and the *star forming* biasing models (see e.g. Marinoni et al. 2005). In the first model the number of galaxies (halo occupation) is conserved as a function of time and the bias evolves assuming the form presented in Eq. 11. The second model takes into account galaxy merging, as proposed by Mo & White (1996), who gave analytical prescriptions for computing the bias of halos using the Press & Schechter formalism. If we assume that galaxies can be identified with dark matter halos, an approximate expression for the biasing of all halos of mass  $> M$  existing at redshift  $z$  (but which collapsed at redshift greater than the observation redshift, see discussion in Matarrese et al. 1997) is given by:

$$b(M, z) = 1 + \frac{1}{\delta_c} \left( \frac{\delta_c^2}{\sigma^2(M, z)} - 1 \right) \quad (22)$$

where  $\delta_c \simeq 1.69$  is the linear overdensity of a sphere which collapses in an Einstein-de Sitter Universe and  $\sigma(M, z)$  is the linear rms fluctuations on scales corresponding to mass  $M$  at the redshift of observation. The *star-forming* model assumes that the distribution of galaxies with luminosity  $> L$  is well traced by halos with mass  $> M$ , and predicts the biasing of objects that just collapsed at the redshift of observation (e.g., Blanton et al. 2000). In this model,

$$b(M, z) = 1 + \frac{\delta_c}{\sigma^2(M, z)} \quad (23)$$

represents the biasing of galaxies that formed in a narrow time interval around redshift  $z$  (i.e., galaxies which experienced recent star formation at redshift  $z$ ).

From the halo mass analysis presented in Section 5.2 we infer that VUDS galaxies would evolve into galaxies with a luminosity  $M_r < -21.5$  in the local universe (Section 5.2), for which the bias is  $b_L = 1.7$ . The star forming model predicts that the VUDS galaxy population would evolve into a weakly biased class of galaxies, corresponding to faint low mass galaxies. The merging model seem to have an evolution too steep

compared to what is observed as it predicts an evolution to a bias significantly lower than populations of galaxies with the lowest clustering at  $z=0$  (Zehavi et al. 2011). The model that would make the VUDS galaxies with  $b_L \sim 2.8$  at  $z=3$  evolve to  $b_L = 1.7$  at  $z=0$  is the passive model. However, it is likely that more sophisticated models would need to be developed to reproduce the joint evolution of DM halo mass and galaxy bias.

## 6. Summary

We examined the galaxy clustering properties of high redshift galaxies with  $2 < z < 5$ , by measuring the two-point real-space correlation function  $w_p(r_p)$  for 3022 galaxies from the VUDS survey. These measurements are among the first performed from spectroscopically confirmed galaxies at such high redshifts and allow to probe clustering evolution and galaxy host halos since very early times. We quantify our observations in terms of a power-law approximation and a three-parameter HOD model fitted to the correlation function computed in three redshift ranges, with median  $z \sim 2.5$ ,  $z \sim 3.0$  and  $z \sim 3.5$ . Our work therefore complements and extends to higher redshifts previous galaxy clustering analyses.

The results and conclusions can be summarized as follows:

- We observe a mild clustering evolution for the star-forming population in the VUDS survey. The correlation length increases from  $r_0 = 3.95 \pm 0.48$  at  $z \sim 2.5$  to reach  $r_0 = 4.35 \pm 0.60$  at  $z = 3.5$ . We attribute this slight increase to the difference in absolute magnitudes between these two samples, the higher redshift sample being more luminous. The slope of the correlation function is found to be  $\gamma = 1.81^{+0.02}_{-0.06}$  and  $1.60^{+0.12}_{-0.13}$  at  $z = 2.5$  and  $3.5$  respectively, similar or slightly flatter than the canonical  $\gamma = 1.8$  observed at low redshifts. These measurements are in general agreement with the most recent and least biased measurements available in the literature at similar redshifts. We find that the power-law parameters we derive from the correlation function measurements are comparable to the clustering parameters for star-forming galaxies at intermediate redshifts, but the clustering length is somewhat lower than for galaxies in the local Universe.
- The average halo mass obtained from our HOD model fit reaches  $\langle M_h \rangle = 10^{11.75} h^{-1} M_\odot$  at redshift  $z = 3$ , which is smaller than observed locally. Assuming the mass growth model from van den Bosch (2002) these halos should evolve into halos with a mass  $\langle M_h \rangle(z = 0) = 10^{13.9} h^{-1} M_\odot$  at  $z = 0$ . As halos with such a mass are occupied by star forming galaxies with  $M_r < -21.5$  in the local universe, we infer that the star-forming galaxy population observed at a redshift  $z \sim 3$  has evolved into the massive and bright galaxy population observed today, as predicted in the hierarchical mass growth paradigm.
- We found that the linear large-scale galaxy bias  $b_L$ , computed using both the HOD and power-law best-fit parameters, increases with redshift from  $b_L^{PL} = 2.68 \pm 0.22$  at  $z = 2.5$  to  $b_L^{PL} = 3.26 \pm 0.47$  at  $z = 3.5$ . These bias values are comparable with measurements obtained at similar redshifts but are much higher than bias values observed at low and intermediate redshift ranges where  $b_L$  reaches values in the range 1 – 2. We discuss simple models to link the  $z \sim 3$  galaxies in our sample to lower redshift popula-

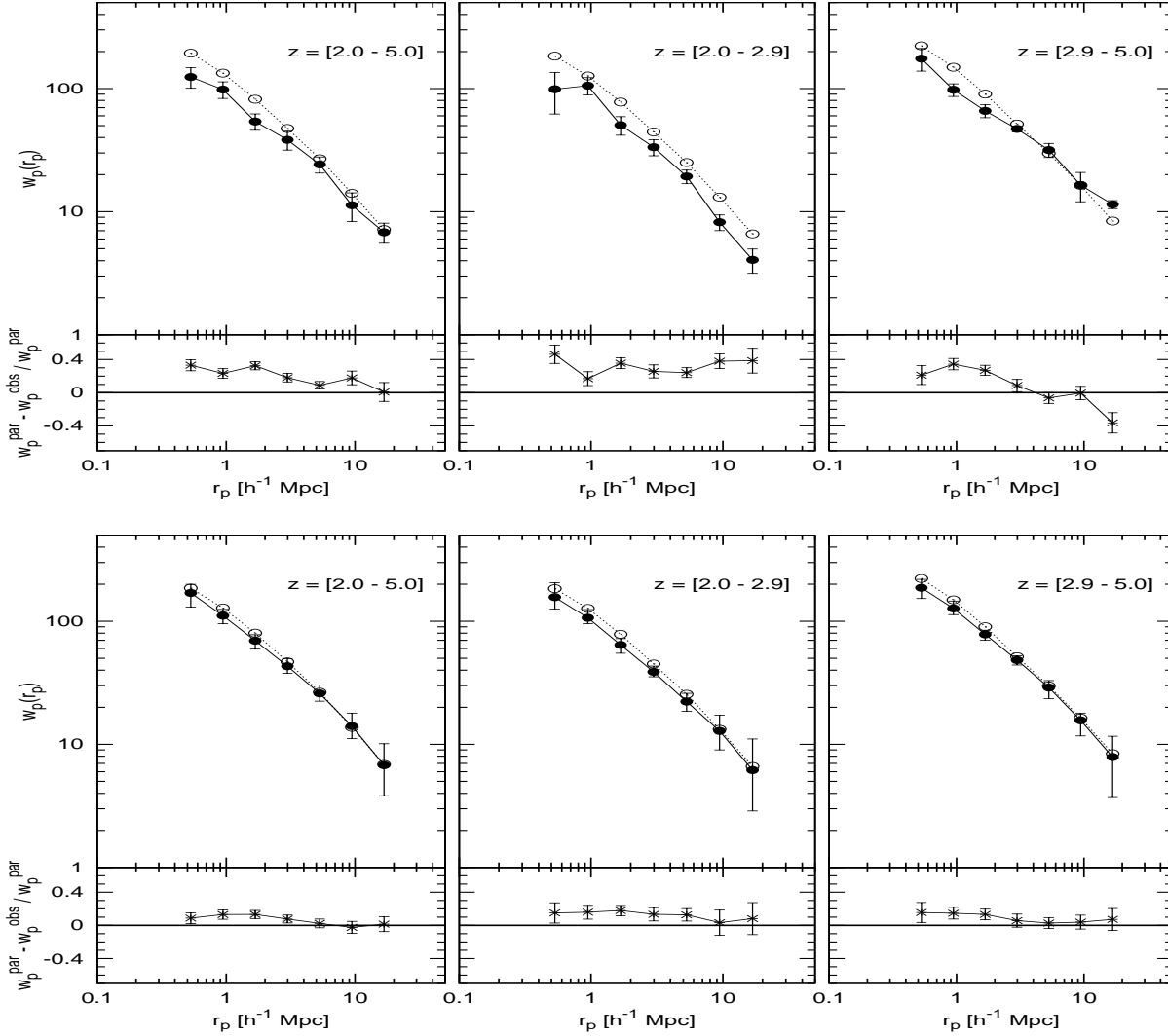
tions, and find that a passive halo evolution model maintaining halo occupation parameters constant will lead to a bias value close to  $b_L = 1.7$  at  $z \sim 0$  we inferred from the halo mass evolution. However it is likely that more sophisticated models are needed to fully interpret this evolution.

Our data provide a robust reference for the clustering of galaxies at an epoch before the peak in star formation rate. We infer that the star-forming population of galaxies at  $z \sim 3$  should evolve into the massive and bright ( $M_r < -21.5$ ) galaxy population which typically occupies haloes of mass  $\langle M_h \rangle = 10^{13.9}$  at redshift  $z = 0$ . More complete modeling seems necessary to relate high redshift to low redshift galaxy populations, possibly taking into account non-linear effects, to better explain the galaxy clustering properties at early epochs in the life of the universe.

**Acknowledgements.** We thank Martin Kilbinger for assistance with implementing our HOD model into the CosmoPMC code. This work is supported by funding from the European Research Council Advanced Grant ERC-2010-AdG-268107-EARLY and by INAF Grants PRIN 2010, PRIN 2012 and PICS 2013. AC, OC, MT and VS acknowledge the grant MIUR PRIN 2010–2011. This work is based on data products made available at the CESAM data center, Laboratoire d'Astrophysique de Marseille. This work partly uses observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at TERAPIX and the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS.

## References

- Abbas, U., de la Torre, S., Le Fèvre, O., et al. 2010, *MNRAS*, 406, 1306
- Adelberger, K. L., Steidel, C. C., Pettini, M., et al. 2005, *ApJ*, 619, 697
- Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, *MNRAS*, 310, 540
- Barone-Nugent, R. L., Trenti, M., Wyithe, J. S. B., et al. 2014, *ApJ*, 793, 17
- Barrow, J. D., Bhavsar, S. P., & Sonoda, D. H. 1984, *MNRAS*, 210, 19P
- Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, *Phys. Rep.*, 367, 1
- Bielby, R., Hill, M. D., Shanks, T., et al. 2013, *MNRAS*, 430, 425
- Blake, C., Collister, A., & Lahav, O. 2008, *MNRAS*, 385, 1257
- Blanc, G. A., Lira, P., Barrientos, L. F., et al. 2008, *ApJ*, 681, 1099
- Blanton, M., Cen, R., Ostriker, J. P., & Strauss, M. A. 1999, *ApJ*, 522, 590
- Blanton, M., Cen, R., Ostriker, J. P., Strauss, M. A., & Tegmark, M. 2000, *ApJ*, 531, 1
- Bottini, D., Garilli, B., Maccagni, D., et al. 2005, *PASP*, 117, 996
- Brown, M. J. I., Zheng, Z., White, M., et al. 2008, *ApJ*, 682, 937
- Cappé, O. & Moulines, E. 2007, *ArXiv e-prints*
- Carroll, S. M., Press, W. H., & Turner, E. L. 1992, *ARA&A*, 30, 499
- Coil, A. L., Newman, J. A., Cooper, M. C., et al. 2006, *ApJ*, 644, 671
- Coil, A. L., Newman, J. A., Croton, D., et al. 2008, *ApJ*, 672, 153
- Connolly, A. J., Scranton, R., Johnston, D., et al. 2002, *ApJ*, 579, 42
- Cooray, A. & Sheth, R. 2002, *Phys. Rep.*, 372, 1
- Coupon, J., Kilbinger, M., McCracken, H. J., et al. 2012, *A&A*, 542, A5
- Daddi, E., Röttgering, H. J. A., Labbé, I., et al. 2003, *ApJ*, 588, 50
- Davis, M. & Peebles, P. J. E. 1983, *ApJ*, 267, 465
- de la Torre, S., Guzzo, L., Peacock, J. A., et al. 2013, *A&A*, 557, A54
- de la Torre, S., Le Fèvre, O., Arnouts, S., et al. 2007, *A&A*, 475, 443
- de la Torre, S., Le Fèvre, O., Porciani, C., et al. 2011a, *MNRAS*, 412, 825
- de la Torre, S., Meneux, B., De Lucia, G., et al. 2011b, *A&A*, 525, A125
- De Lucia, G., Springel, V., White, S. D. M., Croton, D., & Kauffmann, G. 2006, *MNRAS*, 366, 499
- Fisher, K. B., Davis, M., Strauss, M. A., Yahil, A., & Huchra, J. P. 1994, *MNRAS*, 267, 927
- Foucaud, S., McCracken, H. J., Le Fèvre, O., et al. 2003, *A&A*, 409, 835
- Garilli, B., Fumana, M., Franzetti, P., et al. 2010, *PASP*, 122, 827
- Geach, J. E., Sobral, D., Hickox, R. C., et al. 2012, *MNRAS*, 426, 679
- Giocoli, C., Tormen, G., Sheth, R. K., & van den Bosch, F. C. 2010, *MNRAS*, 404, 502
- Guzzo, L., Scodeggio, M., Garilli, B., et al. 2014, *A&A*, 566, A108
- Guzzo, L., Strauss, M. A., Fisher, K. B., Giovanelli, R., & Haynes, M. P. 1997, *ApJ*, 489, 37
- Hamilton, A. J. S. 1993, *ApJ*, 417, 19
- Hartley, W. G., Almaini, O., Cirasuolo, M., et al. 2010, *MNRAS*, 407, 1212
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, 457, 841
- Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, *A&A*, 556, A55
- Iovino, A., Cucciati, O., Scodeggio, M., et al. 2010, *A&A*, 509, A40
- Kaiser, N. 1987, *MNRAS*, 227, 1
- Kashikawa, N., Yoshida, M., Shimasaku, K., et al. 2006, *ApJ*, 637, 631
- Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999, *MNRAS*, 303, 188
- Kilbinger, M., Benabed, K., Cappe, O., et al. 2011, *ArXiv e-prints*
- Kim, J.-W., Edge, A. C., Wake, D. A., et al. 2014, *MNRAS*, 438, 825
- Kong, X., Daddi, E., Arimoto, N., et al. 2006, *ApJ*, 638, 72
- Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., et al. 2004, *ApJ*, 609, 35
- Landy, S. D. & Szalay, A. S. 1993, *ApJ*, 412, 64
- Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, *A&A*, 559, A14
- Le Fèvre, O., Guzzo, L., Meneux, B., et al. 2005, *A&A*, 439, 877
- Le Fèvre, O., Saisse, M., Mancini, D., et al. 2003, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 1670–1681
- Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2014, *ArXiv e-prints*
- Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, *ApJS*, 172, 70
- Lin, L., Dickinson, M., Jian, H.-Y., et al. 2012, *ApJ*, 756, 71
- Magliocchetti, M. & Porciani, C. 2003, *MNRAS*, 346, 186
- Marinoni, C., Le Fèvre, O., Meneux, B., et al. 2005, *A&A*, 442, 801
- Marulli, F., Bolzonella, M., Branchini, E., et al. 2013, *A&A*, 557, A17
- Matarrese, S., Coles, P., Lucchin, F., & Moscardini, L. 1997, *MNRAS*, 286, 115
- McCracken, H. J., Capak, P., Salvato, M., et al. 2010, *ApJ*, 708, 202
- Meneux, B., Guzzo, L., de la Torre, S., et al. 2009, *A&A*, 505, 463
- Meneux, B., Guzzo, L., Garilli, B., et al. 2008, *A&A*, 478, 299
- Meneux, B., Le Fèvre, O., Guzzo, L., et al. 2006, *A&A*, 452, 387
- Mo, H. J. & White, S. D. M. 1996, *MNRAS*, 282, 347
- Mo, H. J. & White, S. D. M. 2002, *MNRAS*, 336, 112
- Monaco, P., Theuns, T., & Taffoni, G. 2002, *MNRAS*, 331, 587
- Moster, B. P., Naab, T., & White, S. D. M. 2013, *MNRAS*, 428, 3121
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *ApJ*, 490, 493
- Norberg, P., Baugh, C. M., Hawkins, E., et al. 2002, *MNRAS*, 332, 827
- Ouchi, M., Shimasaku, K., Okamura, S., et al. 2004, *ApJ*, 611, 685
- Peacock, J. A. & Smith, R. E. 2000, *MNRAS*, 318, 1144
- Peebles, P. J. E. 1980, *The Large-Scale Structure of the Universe* (Princeton: Princeton Univ. Press)
- Percival, W. J., Reid, B. A., Eisenstein, D. J., et al. 2010, *MNRAS*, 401, 2148
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013, *ArXiv e-prints*
- Pollo, A., Guzzo, L., Le Fèvre, O., et al. 2006, *A&A*, 451, 409
- Pollo, A., Meneux, B., Guzzo, L., et al. 2005, *A&A*, 439, 887
- Prada, F., Klypin, A. A., Cuesta, A. J., Betancort-Rijo, J. E., & Primack, J. 2012, *MNRAS*, 423, 3018
- Press, W. H. & Schechter, P. 1974, *ApJ*, 187, 425
- Quadri, R. F., Williams, R. J., Lee, K.-S., et al. 2008, *ApJ*, 685, L1
- Ross, A. J. & Brunner, R. J. 2009, *MNRAS*, 397, 878
- Savoy, J., Sawicki, M., Thompson, D., & Sato, T. 2011, *ApJ*, 737, 92
- Sawangwit, U., Shanks, T., Abdalla, F. B., et al. 2011, *MNRAS*, 416, 3033
- Schneider, P. 2006, *Extragalactic Astronomy and Cosmology*
- Seljak, U. 2000, *MNRAS*, 318, 203
- Skibba, R. A., Bamford, S. P., Nichol, R. C., et al. 2009, *MNRAS*, 399, 966
- Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, *ApJ*, 688, 709
- Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, *ApJ*, 724, 878
- Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005, *ApJ*, 631, 41
- van den Bosch, F. C. 2002, *MNRAS*, 331, 98
- Wake, D. A., Sheth, R. K., Nichol, R. C., et al. 2008, *MNRAS*, 387, 1045
- White, S. D. M. & Rees, M. J. 1978, *MNRAS*, 183, 341
- Wraith, D., Kilbinger, M., Benabed, K., et al. 2009, *Phys. Rev. D*, 80, 023507
- Zehavi, I., Eisenstein, D. J., Nichol, R. C., et al. 2005a, *ApJ*, 621, 22
- Zehavi, I., Weinberg, D. H., Zheng, Z., et al. 2004, *ApJ*, 608, 16
- Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, *ApJ*, 736, 59
- Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005b, *ApJ*, 630, 1
- Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, *ApJ*, 633, 791



**Fig. 1.** Projected two point correlation function  $w_p(r_p)$  measured for 66 VUDS mock catalogs in three redshift ranges. The points correspond to the mean measurement of all 66 mock catalogs, while the errors are computed as their standard deviations. The 'true'  $w_p(r_p)$  computed for the whole parent sample (open circles) is compared to that measured from the 'observed' sample (filled circles). *Upper panel:* Without corrections (2) and (3). *Lower panel:* After our full correction scheme was applied.

## Appendix A: Correlation function corrections scheme

In the galaxy surveys, there exists a number of observational biases which may influence the clustering measurements. These biases are related to the construction of the survey. Both the parent photometric catalogue, from which target galaxies are selected, and the final spectroscopic catalogue may be the source of these biases. Below, we list the most important survey features that cause the biases. Next, we describe the corrections applied to minimize their effects. Since the VUDS survey is performed on the same multi-slit spectrograph VIMOS as the VVDS, zCOSMOS and VIPERS surveys, the correction scheme adopted here is based on the same methodology and the ones used for the other VIMOS-based surveys in the past. In particular, the main corrections used here have been proposed and fully described by Pollo et al. (2005) and de la Torre et al. (2011a).

The main observational biases in the VUDS data can be listed as follows.

1. As shown in Figure 1, some of the areas are excised from the observations, due to the VIMOS lay-out (see A.1). The field of view of this instrument consists of four quadrants separated by  $2'$  gaps. No galaxies are observed in these gaps between quadrants, which influences the pair counts.
2. Obviously, not all galaxies from the photometric target candidate sample can be spectroscopically observed. Each galaxy spectrum occupies a certain area on the CCD detector, calculated from the spatial extent of the slit and the length of the spectrum due to spectral dispersion, which imposes geometrical constraints in target selection as galaxies too close to each other cannot be targeted simultaneously. It means that the galaxies from the parent photometric sample must be chosen in a specific way to effectively allocate as many as possible 'spectral slits' for each observation.

In the case of the VIMOS observations, the slit allocation is performed automatically by the Super-SPOC code (Bottini et al. 2005). As a result, the spectroscopically observed sample is not a random representation of the parent photometric sample, and the introduced bias is relatively complex, especially on the small scales.

These biases, as well as a few other, less important, effects are described in detail by Pollo et al. (2005).

However, these biases can be minimized by a combination of corrections, which were presented by Pollo et al. (2005) and de la Torre et al. (2011a), and which we found the best working for the VUDS data:

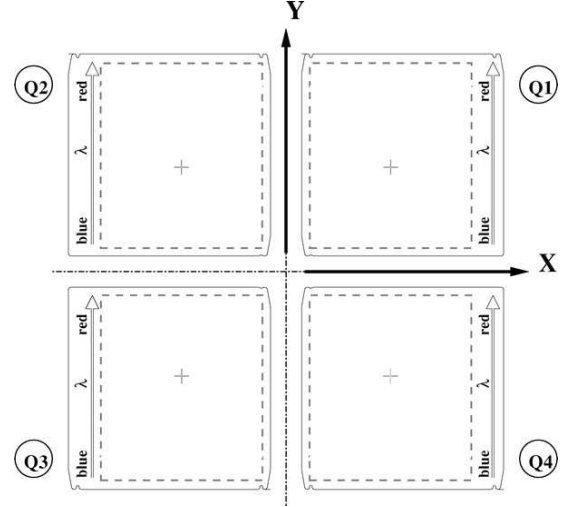
1. Random sample construction. The first, most basic part of the correction scheme is the appropriate construction of the random catalog which needs to be geometrically identical (with the exception of the small scale non-linear biases introduced by the SSPOC) with the spectroscopic sample. Generating random objects in this catalogue, we take into account the shape of the single pointings and quadrants, which are related to the shape of the VIMOS spectrograph. Additionally, we exclude the regions removed from the parent photometric sample (e.g. due to the presence of a bright star), by applying the same photometric mask to the random sample. These first-order corrections reduce most of the negative effects on the correlation function.
2. A global correction. In order to account for the missing pairs (due to the VIMOS limitations and the SSPOC strategy) we assign a "global" weight to each galaxy-galaxy pair. Assuming that the parent catalogue is free from angular incompleteness, we define a weighting function  $f(\theta)$  as a ratio between the mean number of pairs in the parent photometric catalogue and the main number of pairs in the spectroscopic sample, as a function of angular separation (de la Torre et al. 2011a):

$$f(\theta) = \frac{1 + w_{par}(\theta)}{1 + w_{spec}(\theta)}, \quad (\text{A.1})$$

where  $w_{par}(\theta)$  and  $w_{spec}(\theta)$  are the angular correlation functions of the parent photometric and spectroscopic samples, respectively. Then, each pair from the spectroscopic sample separated by the angular distance  $\theta$  is weighted by this ratio  $f(\theta)$ .

3. Small scale corrections. As it comes out, even this strategy does not fully account for the small-scale angular effects of the SSPOC target selection strategy. In order to account for these local small-scale biases we additionally use the local weighting scheme which is also using the parent photometric catalog as a reference. In this case, we count how many galaxies around the targeted galaxy we are missing due to the limited space for spectroscopic slits. Each targeted galaxy pair is then weighted proportionally to its representativeness of the surrounding density field. Similarly to Pollo et al. (2005) we found that the optimal size of the weighting area is the circle with a radius  $\sim 40''$ .

The correcting scheme including all these ingredients was confirmed to be optimal after a series of tests performed using all 66 VUDS mock catalogues. In figure .1 we present the results of the correlation function measurements for the mock



**Fig. A.1.** Lay-out of the VIMOS field of view. Picture from Bottini et al. (2005).

catalogues in our three redshift ranges. All the points correspond to the means of the results from 66 VUDS mock catalogues, while the errors are computed as the standard deviations between them. The three upper panels show the comparison between (1) the 'true' (computed for whole mock parent sample) correlation function and (2) the 'observed' correlation function computed only with the most obvious correction number (1), i.e. the appropriate geometrically cut random sample. The lower panels present the comparison of the same 'true' correlation function and (3) the 'observed' correlation function when the full correcting scheme described above was applied.

It is clearly seen that the introduced corrections significantly improve the retrieved signal from the 'observed' sample at all the spatial scales, and in all three redshift bins considered.

However, it can be noticed that on small scales ( $r_p < 3h^{-1}\text{Mpc}$ ), there is still some signal missing, even after introducing the full optimal correcting scheme (see lower panels of Figure .1). For these separations, our measurement is on average underestimated by  $10\% \pm 3\%$  with respect to the 'true' value of  $w_p(r_p)$  computed from the parent sample. The possible influence of this effect on the final results is discussed in section 3.3. At separations larger than  $r_p \sim 3h^{-1}\text{Mpc}$  the value of  $w_p(r_p)$  is also slightly underestimated, but only by about  $1.6\% \pm 0.9\%$ .

Our correction scheme provides a significant improvement in comparison to the average 28% bias observed without applying it. However, the remaining systematic underestimation of the  $w_p(r_p)$  has to be taken into account during the correlation function measurements (see Sec. 3.3).

## Appendix B: Projected correlation function measurements

For every galaxy sub-sample used in this work we present the tables of the projected correlation function measurements  $w_p(r_p)$  with associated  $1\sigma$  errors at different separations  $r_p$  (in units  $h^{-1}\text{Mpc}$ ).

**Table B.1.** General galaxy population. Projected correlation function measurements of redshift sub-samples. The first column provides the pair-weighted-projected separation of the bin  $r_p$ . Subsequent columns provide the projected correlation function values  $w_p(r_p)$  along with  $1\sigma$  errors.

$r_p$	$w_p(r_p)$		
	$2.0 < z < 2.9$	$2.0 < z < 5.0$	$2.9 < z < 5.0$
0.29	$147.40 \pm 42.02$	$107.65 \pm 32.32$	$68.74 \pm 48.22$
0.53	$84.84 \pm 26.39$	$65.03 \pm 21.72$	$67.51 \pm 36.77$
0.94	$46.83 \pm 12.79$	$42.22 \pm 10.26$	$48.77 \pm 15.83$
1.68	$28.71 \pm 7.57$	$28.11 \pm 6.07$	$41.94 \pm 9.80$
2.98	$22.81 \pm 4.87$	$19.20 \pm 4.50$	$20.21 \pm 6.58$
5.31	$12.32 \pm 4.01$	$12.99 \pm 3.06$	$17.87 \pm 4.13$
9.44	$8.63 \pm 2.39$	$8.97 \pm 2.01$	$10.62 \pm 2.82$
16.79	$4.26 \pm 2.08$	$6.23 \pm 1.62$	$9.44 \pm 2.19$